(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
6 September 2002 (06.09.2002)

PCT

(10) International Publication Number
WO 02/068579 A2

(51) International Patent Classification[7]: C12G

(21) International Application Number: PCT/US02/00284

(22) International Filing Date: 10 January 2002 (10.01.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/756,696 10 January 2001 (10.01.2001) US

(71) Applicant: **PE CORPORATION (NY)** [US/US]; 761 Main Avenue, Norwalk, CT 06859 (US).

(72) Inventors: **VENTER, Craig, J.**; Celera Genomics, 45 West Gude Drive C2-4#21, Rockville, MD 20850 (US). **ADAMS, Mark**; Celera Genomics, 45 West Gude Drive C2-4#21, Rockville, MD 20850 (US). **LI, Peter, W., D.**; Celera Genomics, 45 West Gude Drive C2-4#21, Rockville, MD 20850 (US). **MYERS, Eugene, W.**; Celera Genomics, 45 West Gude Drive C2-4#21, Rockville, MD 20850 (US).

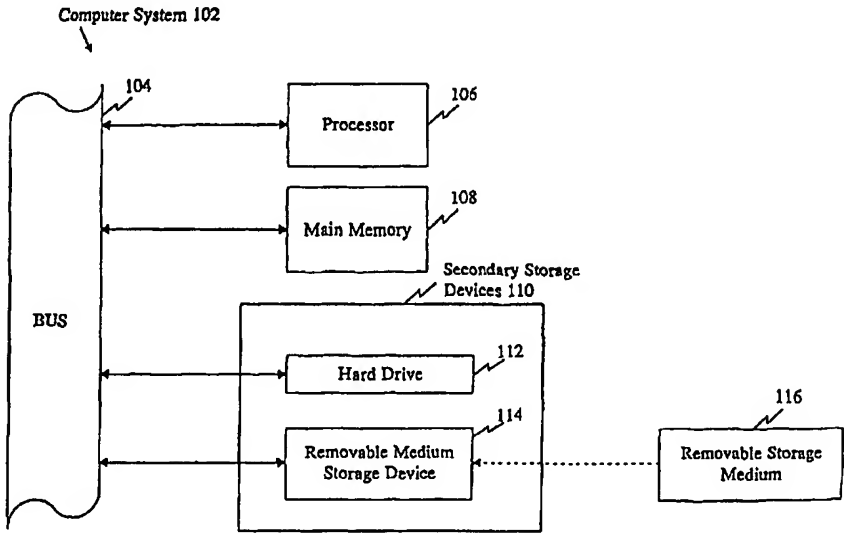
(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: KITS, SUCH AS NUCLEIC ACID ARRAYS, COMPRISING A MAJORITY OF HUMAN EXONS OR TRANSCRIPTS, FOR DETECTING EXPRESSION AND OTHER USES THEREOF



(57) Abstract: The present invention is based on the sequencing and assembly of the human genome. The present invention provides the primary nucleotide sequence of the coding portion of the human genome in the form of a series of transcript sequences with accompanying exon information. This information can be used to generate nucleic acid detection reagents and kits such as nucleic acid arrays, and for other uses.

WO 02/068579 A2

KITS, SUCH AS NUCLEIC ACID ARRAYS, COMPRISING A MAJORITY OF HUMAN EXONS OR TRANSCRIPTS, FOR DETECTING EXPRESSION AND OTHER USES THEREOF

FIELD OF THE INVENTION

The present invention is in the field of genomic discovery systems. The present invention specifically provides the coding sequences of the human genome, including transcript sequences and corresponding exon information, in a form that is commercially useful, including detection kits and reagents such as nucleic acid arrays.

BACKGROUND OF THE INVENTION

The human genome is organized into discrete expression units called genes. Genes are further divided into exons (coding sequences) and introns (intervening, non-coding sequences). RNA transcripts are the primary output of the genome and are generated through a process referred to as gene expression or transcription. Gene expression involves the transcription of DNA into pre-mRNA, followed by RNA processing of pre-mRNA into mature mRNA transcripts, during which introns are removed and exons are spliced together to form complete transcript sequences. However, alternative splicing pathways allow introns to be removed and exons to be combined in different combinations, thereby allowing different mRNAs and proteins to be produced from the same gene. It has been found that nearly 40% of human genes are alternatively spliced (Brett *et al.*, 2000, *FEBS Lett.*, 474, 83). Different splice forms of genes may play distinctly different, and important, roles in different cells/tissues, developmental stages, or diseases and, therefore, the ability to detect different splice forms of the same gene is of paramount importance. Alternative splicing can also act as an on-off mechanism for mRNA activity by producing either functional or non-functional mRNAs from the same pre-mRNA.

A major goal in the development of therapeutics, diagnostic reagents, and pharmaceutical drugs is to understand and elucidate gene expression patterns and splicing patterns, particularly in different cells/tissues, developmental stages, and disease/pathological conditions. Determining when or under what conditions a particular gene or splice form is expressed, in which cells/tissues, and to what extent is important for understanding the function of the protein encoded by the gene and its role in disease. Gene expression and splicing patterns can be determined by reagents or kits, preferably nucleic acid arrays (also known as "DNA chips" or "biochips"), that utilize detection elements, such as nucleic acid probes, to detect the expression of gene fragments or the splicing together of exons to form mRNA transcripts. Such detection

elements may comprise, for example, fragments of, or complete, gene transcripts or exons, fragments corresponding to UTR regions of the transcript or detection elements that span the exon/exon boundaries of a transcript. The use of exons, or exon fragments, as detection elements has the distinct advantage of allowing the detection of different alternatively spliced transcript forms with the same detection element. This is possible so long as the transcript form contains the particular exon that is used as a detection element, regardless of how that exon is combined with other exons. On the other hand, the use of complete transcripts, or transcripts comprising more than one exon, generally allows the detection of only that particular splice form, or exon combination, and may not detect the expression of other important transcript splice forms. However, the use of transcripts as detection elements is advantageous in particular situations, such as when detection of only one particular transcript, with a high degree of specificity and minimal cross hybridization to other transcript forms, is desired.

The primary sequence of the exons/transcripts of the human genome would therefore be valuable for use in detection kits and reagents, such as nucleic acid arrays, for detecting gene expression patterns, including variable gene expression such as alternative splicing, and other uses. Human exons/transcripts can serve as detection elements, such as probes, in detection kits and reagents such as nucleic acid arrays. Not only will such kits and reagents serve as a basis for discovery and validation of commercially important genes, they provide commercially valuable tools for understanding the complex patterns of gene expression in relationship to different cells/tissues, developmental stages, and disease conditions. Consequently, human exons/transcripts, provided in a usable form, such as in the form of detection elements in a nucleic acid array, would be valuable for disease diagnosis and treatment, such as by improving the drug discovery and development process, or for diagnosing diseases based on aberrant gene expression patterns.

Furthermore, a substantial proportion of current gene discovery efforts is directed at mining EST databases. However, it has been estimated that EST databases may contain as little as 40% of the protein-coding portion of the human genome (Aparicio, *Nature Genetics*, June 2000, 25: 129-130). Consequently, the primary sequence of human transcripts and exons, identified through whole-genome sequencing, assembly, and annotation, represents the best source of identifying protein-coding sequences of the human genome that are not represented in EST databases. Therefore, the sequence of human exons/transcripts provided by the present invention is useful for identifying and validating commercially valuable human genes.

Gene expression analysis, using the transcript/exon sequences provided herein, is also useful for determining functions and relationships of genes with unknown functions. For

example, it has been shown in yeast that genes with similar functions have similar gene expression profiles (Eisen *et al.* (1998) *Proc. Natl. Acad. Sci. U.S.A.* 95, 14863-14868).

The present invention advances the art by providing the predicted transcript sequences (SEQ ID NOS:1-39010), for 39010 transcripts predicted from the assembled human genome, many of which did not have evidence for their existence in the prior art. Furthermore, the present invention provides information on each of the exons (Table 1) contained within the transcripts. The exon information contained in Table 1 includes the coordinates of each exon within its respective transcript, thereby allowing one to readily determine the precise boundaries of each of the exons using the transcript coordinates and the transcript sequences as a reference. These exon boundaries define the exon-exon junctions discussed herein. Also provided in Table 1 is evidence supporting the existence of each exon or transcript (e.g. EST hit, mouse hit, etc.).

Given the transcript sequences provided by the present invention and the exon coordinate information provided in Table 1, or fragments thereof, readily implementable compositions of matter, such as detection elements and detection reagent/kits, (e.g. in the form of probes in a nucleic acid array), can be made using methods well known in the art and discussed herein. Such kits and reagents can be used to track the expression and/or splicing of all of the transcripts/genes disclosed herein, the novel members herein provided, or rationally selected subsets thereof, defined by a user.

Nucleic Acid Arrays and Detection Kits and Reagents

Oligonucleotide probes have long been used to detect complementary nucleic acid sequences in a nucleic acid of interest (the "target" nucleic acid) in the form of detection kits and reagents. In some assay formats, the oligonucleotide probe is tethered, i.e., by covalent attachment, to a solid support, and arrays of oligonucleotide probes immobilized on solid supports have been used to detect specific nucleic acid sequences in a target nucleic acid. See, e.g., PCT patent publication Nos. WO 89/10977 and 89/11548. In other formats, the detection reagents are supplied in solution.

The development of arraying technologies such as photolithographic synthesis of a nucleic acid array and high density spotting of cDNA products has provided methods for making very large arrays of oligonucleotide probes in very small areas. See U.S. Pat. No. 5,143,854 and PCT patent publication Nos. WO 90/15070 and 92/10092. Microfabricated arrays of large numbers of oligonucleotide probes, called "DNA chips" offer great promise for a wide variety of applications. Such arrays may contain, for example, thousands or millions of probes. Probes may

be formed from, for example, cDNA clones, PCR products, or oligonucleotides and can be used in solution or tethered to a support such as a planar surface (chip) or bead format.

The present invention provides detection kits and reagents, such as nucleic acid arrays, that are based on the novel transcript/exon sequences of the human genome provided herein, particularly the novel transcripts and novel information concerning exon structure of each transcript provided in the Sequence Listing and in Table 1.

Medical Importance of Variable Gene Expression

Variable gene expression, such as alternative splicing (also referred to by such terms as alternate splicing or differential splicing) and alternative start/termination sites, is a fundamentally important mechanism of gene regulation. Alternative splicing refers to the formation of two or more different mature mRNA splice forms from a single gene or pre-mRNA, depending on the combination of exons that are spliced together. Alternative splicing therefore serves as an important means of generating additional protein diversity from the structural information encoded by genes. Furthermore, expression of particular splice forms may differ between, for example, cells, tissues, developmental stages/ages, populations or sexes, and may be altered in certain disease states, such as cancer. Alternative splicing may have a detrimental effect on intercellular interactions and the interaction of various polypeptides and cytokines and thereby lead to diseases such as cancer.

Detection reagents, such as nucleic acid arrays and other multi transcript detection reagent/kit, that utilize detection elements comprised of individual transcripts or exons are capable of detecting alternative splice forms of genes that may be missed by detection reagents that detect only one transcript form. Detection reagents that detect disease-specific splice forms of a gene are useful for disease diagnosis. For example, one or more detection reagents to each exon can be used to determine if an exon is present in a sample and/or detection reagents that span exon/exon boundaries can be used to see if a particular exon/exon splice junction is present and also selects against cross reactivity with genomic DNA.

Alternative splicing plays an important role in a variety of proteins and disease pathways, as the following examples illustrate. Elastin is a protein that is important for providing the elastic properties of the lungs, large blood vessels, and skin. The primary elastin transcript undergoes substantial alternative splicing, and it has been suggested that such alternative splicing of elastin may be population-specific and contribute to aging and pathological conditions in the cardiovascular and pulmonary systems (Indik *et al., Am J Med Genet* 1989 Sep;34(1):81-90).

Nitric oxide proteins are important in numerous physiological processes, such as neurotransmission and muscle relaxation. At least six different isoforms of neuronal nitric oxide mRNA have been identified and found to differ in enzymatic properties. Alternative splicing provides a mechanism to generate this diversity. Furthermore, it has been observed that an alternatively spliced form of neuronal nitric oxide that lacks exon 2 is expressed in many human brain tumors (Brenman *et al., Dev Neurosci* 1997;19(3):224-31).

Alternative splicing of the amyloid precursor protein mRNA, particularly variant splicing of exons 7 and 15, may be involved in the development of Alzheimer's disease (Beyreuther *et al., Ann N Y Acad Sci* 1993 Sep 24;695:91-102).

A number of different estrogen receptor mRNA variants, many of which are generated by alternative splicing, have been identified in breast cancer tissue and may be associated with the development and progression of breast cancer (McGuire *et al., Mol Endocrinol* 1991 Nov;5(11):1571-7).

CD44 is a large family of transmembrane glycoprotein isoforms that are generated from a single gene by alternative splicing and are involved in a variety of cancers. For example, some CD44 isoforms have been found to be causally involved in lung metastasis formation. Furthermore, the expression levels of particular CD44 isoforms are indicative of prognosis in numerous cancers, such as non-Hodgkin lymphomas; gastric, colon, renal, and mammary carcinomas; and in neuroblastomas (Gunthert *et al., Cancer Surv* 1995;24:19-42 and Ponta *et al., Invasion Metastasis* 1994-95;14(1-6):82-6). Therefore, detecting the expression of CD44 alternative splice forms is useful for diagnosing diseases such as these cancers.

Alternative splicing at three positions on the primary fibronectin transcript generates multiple fibronectin polypeptide variants. Furthermore, these different fibronectin variants play specific roles in fibronectin dimer secretion, blood clotting, adhesion to lymphoid cells, skin wound healing, atherosclerosis, and liver fibrosis (Kornblihtt *et al., FASEB J* 1996 Feb;10(2):248-57).

Alternative splicing is important in the differentiation, maintenance, and function of the red blood cell membrane. This is highlighted by the finding that hereditary hemolytic anemias result from mutations that cause defective splicing (Benz *et al., Trans Am Clin Climatol Assoc* 1996;108:78-95).

Platelet derived growth factor (PDGF), which is associated with several diseases including atherosclerosis and neoplasia, undergoes alternative splicing that could affect the function of PDGF (Khachigian *et al., Pathology* 1992 Oct;24(4):280-90); consequently,

alternative splicing of PDGF may play a significant role in diseases such as atherosclerosis and neoplasia.

As an example of the importance of alternative splicing in development, it is well known in the art that sex-specific alternative splicing in *Drosophila* plays an important role in sex determination.

Additionally, exonic splicing enhancers (ESEs) are sequence elements within exons that promote splicing and it has been suggested that many human diseases linked to mutations or polymorphisms within exons may be caused by the inactivation of ESEs, thereby leading to defective splicing (Blencowe, *Trends Biochem Sci* 2000 Mar;25(3):106-10).

As these examples illustrate, such fields as therapeutic/pharmaceutical drug development and disease diagnosis/treatment would greatly benefit from detection kits and reagents that improve the detection of variable gene expression, such as the detection of alternative splice forms.

Using Transcripts/Exons as Detection Elements to Monitor Variable Gene Expression

The transcript sequences and the corresponding exon structure of the transcript disclosed herein are useful in themselves as probe/primer sequences and in the design of such detection element, such as nucleic acid arrays or other detection kits. Transcript sequences with exon structure is particularly useful for studying variable forms of gene expression, such as the expression of alternative splice forms and alternative start/termination sites. As the above examples illustrate, alternative splice forms play important roles in a variety of disease conditions, such as cancer. The importance of detecting alternative splice forms is further highlighted by the finding that nearly 40% of human genes are alternatively spliced (Brett *et al.*, 2000, *FEBS Lett.*, 474, 83); therefore, 40% of all human genes may express alternative transcript forms that are undetectable by conventional detection reagents that are not capable of detecting alternative splice forms of expressed genes.

Individual exons are capable of detecting alternative splice forms that comprise that particular exon, regardless of the combination in which that exon is spliced together with other exons to form an alternative splice form. Therefore, one or more detection elements directed to each single exon can be used to detect any splice form that includes that particular exon. For example, if exon 2 of a six exon gene is used as a detection element (for example, as a probe in a nucleic acid array), that detection element can detect the mRNA splice form of exon 2 with exons 3 and 4, as well as the alternative mRNA splice form of exon 2 with exons 1, 5, and 6. These two different splice forms may have distinct functional properties and one of the two

splice forms may cause a disease condition, or be diagnostic of a disease condition. Exon-based detection elements, such as nucleic acid array probes, may be formed, for example, from exons directly amplified from genomic DNA or synthesized using the sequences provided herein as a reference.

Alternatively, sequences that span an exon/exon junction (see Table 2) can be used to generate detection reagents that are useful in detecting expression and/or splice formation. Such reagents are particularly useful in that detection signal caused by genomic contamination in the sample is greatly reduced.

In addition to alternative splicing, variable gene expression also includes alternative start and termination sites. As with alternative splicing, detection reagents that employ individual exons are useful for detecting transcripts with alternative start and/or alternative termination sites, so long as the transcript includes the exon that comprises the detection element.

Commonly used detection techniques that utilize one transcript form comprised of multiple exons spliced together in a particular combination, such as probes formed from cDNA libraries, are limited in that they will not detect transcripts that are comprised of exons spliced together in a different combination, even if some of the exons are the same. Furthermore, such detection elements may not detect transcripts that comprise alternative start and/or termination sites. This prevents the detection of particular splice forms that may play important roles in, for example, certain disease pathways. Therefore, in certain applications, exon sequences are preferable to larger transcript sequences.

Accordingly, a definite need exists in the art for exons of the human genome provided in a useful form, such as in the form of detection elements of a nucleic acid array or other detection reagent/kit. Exons provided in such a form would be extremely valuable for detecting alternative splice forms and other forms of variable gene expression.

Using Sequences that Span Exon-Exon Junctions as Detection Elements

Sequences that span exon-exon junctions in a transcript are especially useful as detection elements, such as probes in a nucleic acid array. In particular, sequences that span exon-exon junctions eliminate false signals caused by genomic contamination. This is because a detection element comprising two neighboring exons as one contiguous sequence will not hybridize to genomic DNA comprising intervening intronic DNA. Such detection elements will only hybridize to expressed mRNA transcripts in which the exons are connected and the intronic sequence has been removed, thereby forming one contiguous stretch of sequence corresponding to the sequence of the detection element that spans the exon-exon junction.

Exon-exon junctions are provided by the present invention and identified in Table 1. Sequences spanning exon-exon junctions can readily be determined using the exon coordinates provided in Table 1 along with the transcript sequences provided in the Sequence Listing. These detection reagents alone, or in combination with intra-exon probes, can be used to elucidate the splicing and expression pattern of genes within a variety of tissues and/or treatment protocols.

Using Transcripts as Detection Elements to Monitor Gene Expression

Transcript sequences of the human genome are also useful for monitoring gene expression patterns and, in certain circumstances, may be preferable to individual exons for use as detection elements for detecting gene expression. For example, the use of transcripts may be preferred when the goal is to monitor expression of a particular transcript, or group of transcripts, to the exclusion of all other transcripts, such as alternative splice forms. In this situation, using transcripts as detection elements, rather than individual exons, increases specificity and decreases undesired cross hybridization of the detection elements with alternative splice forms.

Accordingly, a definite need exists in the art for transcripts of the human genome, as well as exons, provided in a useful form, such as in the form of detection elements in a detection reagent/kit, such as in a nucleic acid array. Transcripts provided in such a form are useful for monitoring particular forms of gene expression with a high degree of specificity. Such detection elements can readily be generated using the sequence information provided herein.

SUMMARY OF THE INVENTION

The present invention is based on the sequencing and assembly of the human genome. The present invention provides the primary nucleotide sequence of the coding portions of the human genome in a series of predicted transcript sequences generated from the assembled and annotated human genome (SEQ ID NOS:1-39010). Furthermore, the position of each exon contained within these transcripts is identified in Table 1. Individual exon sequences can readily be determined using the transcript sequences of SEQ ID NOS:1-39010 along with the coordinates of each exon within it's respective transcript, as provided in Table 1. This information can be used to readily generate nucleic acid detection reagents and kits, such as nucleic acid arrays. In particular, detection reagents are provided that comprise at least one detection element, wherein at least one detection element comprises a transcript selected from SEQ ID NOS:1-39010. In preferred embodiments, at least one detection element of the detection reagent comprises an exon identified in Table 1. In other preferred embodiments, the detection reagent is a nucleic acid array and the detection elements may be, for example, probes attached

to the surface of the array. Furthermore, in other preferred embodiments, the detection reagent comprises 10,000 or more detection elements, one or more from each of the novel transcripts/exons disclosed herein.

Detection elements that comprise a transcript sequence or an exon, particularly an exon selected from Table 1, allow one to identify variable forms of gene expression, such as different splice forms of genes containing the exon of the detection element. Variable forms of gene expression, such as alternative splicing, may have important tissue-specific, disease-specific, or development-specific expression patterns. Such variable forms of gene expression may go undetected by conventional detection techniques used in gene expression studies. Detection elements that comprise a transcript, particularly a transcript selected from SEQ ID NOS:1-39010, allow one to monitor the expression of the transcript that comprises the detection element with a high degree of specificity.

Furthermore, a preferred class of detection elements provided by the present invention comprises sequences spanning exon-exon junctions. Preferred sequences span one exon-exon junction, however, sequences may span any number of exon junctions. Sequences that span exon-exon junctions are particularly useful in that they eliminate false signals caused by genomic contamination. Exon-exon junctions are provided by the present invention and identified in Table 1. Sequences spanning exon-exon junctions can readily be determined using the exon coordinates provided in Table 1 along with the transcript sequences provided in the Sequence Listing.

The present invention provides the nucleotide sequences of the coding portion of the human genome, namely predicted transcript sequences and corresponding exon information, in a form that can be used, analyzed, and commercialized for other uses in addition to detection kits and reagents. For example, the present invention provides the nucleic acid sequences as contiguous strings of primary sequences in a form readable by computers, such as recorded on computer readable media, e.g., magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media. The present invention specifically provides a CD-R that comprises this sequence information (in the form of a Sequence Listing, provided in file SEQLIST.TXT on the accompanying CD labeled CL001101CDA). Such compositions are useful for, for example, for virtual northern blot analysis, BLAST searching, discovery and validation of drug targets, and for comparative genomic studies between genomes of different organisms.

The present invention further provides systems, particularly computer-based systems, which contain the primary sequence information of the present invention stored in data storage means. Such systems are designed to identify commercially important fragments of the human genome.

Another embodiment of the present invention is directed to isolated fragments, and collections of fragments, of the human genome. The fragments of the human genome include peptide-coding fragments, such as transcripts and exons. The transcript sequences (SEQ ID NOS:1-39010) are provided in the Sequence Listing, which is provided in file SEQLIST.TXT, and the exon elements that each transcript is comprised of are provided in Table 1, which is provided in file TABLE1.TXT. Both files are provided on the accompanying CD labeled CL001101CDA.

As discussed above, the present invention includes detection reagents and kits, such as nucleic acid arrays and microfluidic devices, that comprise one or more fragments of the human genome of the present invention, particularly transcript sequences and/or isolated exon sequences. The kits, such as arrays, can be used to track the expression of many exons or genes, even all of the exons or genes, or rationally selected subsets thereof, contained in the human genome.

The identification of the coding set of sequences from the human genome will be of great value for a variety of commercial purposes. Many fragments of the human genome will be immediately characterized by similarity searches against protein and nucleic acid databases and by identifying structural motifs present in protein domains and will be of immediate value to researchers and for the production of proteins or to control gene expression. A specific example concerns secreted proteins, ion channels and G-protein coupled receptors. The biological significance of secreted proteins for controlling cell signaling, differentiation and proliferation is well known.

Further, the development of therapeutic proteins and protein targets for human intervention typically involves identifying a protein that can serve as a target for the development of a small molecule modulator. Many classes of proteins are well characterized as suitable pharmaceutical drugs (protein therapeutics or modified forms thereof) and/or drug targets. These include, but are not limited to, secreted proteins, GPCRs and ion channels.

Brief Description of the Files contained on CD labeled CL001101CDA

1) File SEQLIST.TXT provides the Sequence Listing of the transcript sequences of the present invention in text (ASCII) format. The file size is 50.7 MB.

2) File TABLE1.TXT provides Table 1, which gives detailed information on exon structure for each of the transcript sequences in the Sequence Listing. The size of this file is 15.2 MB and is stored in text (ASCII) format.

Brief Description of Table 1

Table 1 gives the results of detailed computer analysis of the human genome. Table 1 provides information on every identified human transcript and exon comprising every gene/coding region of the human genome, as follows:

The SEQ ID NO: of each transcript sequence (corresponding to SEQ ID NOS:1-39010 provided in the file, SEQLIST.TXT), a Celera UID identifying number for each transcript, a Celera CT identifying number for each transcript, numbers corresponding to each predicted exon contained within each transcript, predicted exon boundaries (indicating exon-exon junctions) identified by coordinates within the corresponding transcript, and supporting evidence for the existence of each exon and/or transcript, where available (H = human EST/cDNA support, R = rodent EST/cDNA support, M = mouse genomic support, and P = protein homology).

Brief Description of the Figure

The figure provides a block diagram of a computer system 102 that can be used to implement the computer-based systems of the present invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention is based on the sequencing and assembly of the human genome. In this process, the primary nucleotide sequence of over 30 million nucleic acid fragments, from about 400 to about 600 nucleotides in length, was determined. These fragments were assembled using the Celera Assembler. After assembly, the sequences were analyzed with various computer packages and compared with all external data sources. The result of this analysis was the identification of 39010 predicted protein-coding transcripts contained in the human genome. The present invention provides the nucleic acid sequences of these transcripts (SEQ ID NOS:1-39010), along with corresponding exon information, in a form that can be used, for example, to readily develop nucleic acid detection kits and reagents, such as nucleic acid arrays.

The present invention provides the nucleotide sequences of the coding sequences of the human genome, including transcript sequences (SEQ ID NOS:1-39010) and corresponding exon information (provided in Table 1), in a form that can be readily used, analyzed, and interpreted

by a skilled artisan. In one embodiment, the sequences are provided as contiguous strings of primary sequence information corresponding to the nucleotide sequences provided in SEQ ID NOS:1-39010, and/or the exons identified in Table 1; the delineated nucleotide sequence of each exon can readily be determined using the transcript sequences of SEQ ID NOS:1-39010 along with the coordinates of each exon within it's respective transcript, as provided in Table 1. The exon information is provided in file TABLE1.TXT and the transcript sequences are provided in file SEQLIST.txt; both of these files are provided on the accompanying CD labeled CL001101CDA. The information in these files has many commercially important uses. For example, the transcript/exon sequences and structural information provided herein can be used to generate commercially valuable nucleic acid or peptide fragments, design and develop probes/primers, and to develop detection reagents and kits such as gene expression arrays. Furthermore, the sequence and structural information provided herein is valuable for a wide variety of commercially important computer-based biological analysis, such as virtual northern blot analysis of gene expression, BLAST searching, or comparative genomic analysis of different organisms. Uses such as these enable the identification and validation of commercially important genes and gene products, as well as diagnostic kits, therapeutics agents, and drug targets.

In other embodiments, the sequences of the present invention are represented by a detection reagent/kit that is capable of identifying mRNA sequences that hybridize to any particular exonic or transcript sequence provided herein. In particular, detection reagents and kits are provided that comprise at least one detection element, wherein at least one detection element comprises a transcript selected from SEQ ID NOS:1-39010, or a portion thereof. In preferred embodiments, at least one detection element of the detection reagent comprises an exon specific detection element identified in Table 1. In other preferred embodiments, at least one detection element of the detection reagent spans at least one exon-exon junction; exon-exon junctions are identified in Table 1. Furthermore, in preferred embodiments, the detection reagent/kit is a nucleic acid array and the detection elements may be, for example, probes attached to the surface of the array. Other preferred detection reagents/detection elements include TaqMan probe/primer sets, for monitoring gene expression using the TaqMan 5' nuclease PCR assay. Furthermore, in most of the preferred embodiments, the detection reagent comprises about more than one detection element (sequence) and preferably, 10,000 or more such detection elements. Such detection reagents can be used to track the expression of many genes/transcripts, or transcript processing, even all of the transcripts/genes/exons, or rationally selected subsets thereof, contained in the human genome.

As used herein, "detection elements" correspond to an element, such as a nucleic acid probe, a probe/primer pair, or a binding aptamer, that is capable of selectively binding a transcript or exon sequence provided by the present invention, or a fragment thereof. Such detection elements include, for example, isolated oligonucleotides comprising the transcript/exon sequences provided herein, provided in a format such as in an array or in a TaqMan 5' nuclease PCR assay. Detection elements, such as probes/primers, may be, for example, attached to a solid support (e.g., in arrays) or supplied in solution (e.g., probe/primer sets for enzymatic reactions such as PCR or RT-PCR).

Additionally, "detection elements" also include the transcript/exon sequences and/or structural information provided herein implemented in a computer-based system. For example, the transcript/exon sequences provided herein may be used as detection elements for searching a computer-based database of sequence or expression information, such as for sequence similarity searching, virtual northern blot analysis, BLAST searching, gene discovery/validation, gene functional analysis, or comparative genomic/expression studies between different individuals, species/organisms, or disease conditions.

Furthermore, one of the preferred classes of detection elements provided by the present invention comprises detection elements that span an exon-exon junction in a transcript. Preferred detection elements span one exon-exon junction. However, detection elements may span any number of exon-exon junctions within a transcript. Detection elements that span exon-exon junctions are particularly useful in that they eliminate false signals caused by genomic contamination. Exon-exon junctions are identified in Table 1. Sequences spanning exon-exon junctions can readily be determined using the exon coordinates provided in Table 1 along with the transcript sequences provided in the Sequence Listing. Thus, references herein to exon, transcript, or gene sequences also include sequences spanning one or more exon-exon junctions.

"Detection reagents" and "detection kits" refer to any system or technology platform that utilizes detection elements comprising nucleic acid or peptide sequences/molecules/fragments corresponding to the transcripts/exons of the present invention, as described above. Thus, detection reagents or detection kits may refer to, for example, nucleic acid arrays (which may also be referred to by such terms as "DNA chips", "biochips", or "microarrays"), the TaqMan 5' nuclease PCR assay system and probe/primer sets, or other enzymatic or PCR-based assay systems, solutions of probes and/or primers, compartmentalized kits, dot-blot or reverse dot-blot systems, sequencing systems, microfluidic systems, mass spec systems, and various computer-based systems such as databases of nucleic acid sequences, protein sequences, or expressed sequences.

The term "transcript" is generally used herein to refer to coding or expressed segments of the human genome that comprise a set of one or more exons that form a mature mRNA molecule upon transcription/expression. The term "transcript" is also used herein to refer to the mRNA transcript molecule, as well as the set of exons in genomic DNA that comprise the mRNA transcript molecule. "Transcripts" may also be referred to herein as "genes", and vice versa, in order to refer to coding portions of genes or open reading frames (ORFs) that correspond to the transcript/exon sequences provided herein.

As used herein, a "representative fragment of the nucleotide sequence provided herein" refers to any portion of these sequences that are not presently represented within a publicly available database, or more particularly to a collection of fragments, where at least one of the members of the collection is unknown, or the entire set has never been described in it's entirety.

Those in the art will readily recognize that detection elements that are comprised of nucleic acid molecules may be supplied as double stranded molecules and that reference to a particular sequence on one strand refers, as well, to the corresponding complementary sequence on the opposite strand. Thus reference to an adenine, a thymine (uridine), a cytosine, or a guanine on one strand of a nucleic acid molecule is also intended to include the thymine (uridine), adenine, guanine, or cytosine, respectively, at the corresponding sites on a complementary strand of the nucleic acid molecule. Thus, reference may be made to either strand in order to refer to a particular nucleic acid sequence or detection element. Oligonucleotide, such as probes and primers, may be based on, or hybridize to, either strand. Throughout the text, reference is generally made to the protein-coding strand, only for the purpose of convenience.

The nucleotide sequence information provided herein was obtained by sequencing the human genome using a shotgun sequencing method known in the art. The nucleotide sequences provided herein are highly accurate, although not necessarily a 100% perfect, representation of the set of exonic nucleotide sequences of the human genome.

Using the information provided herein together with routine cloning and sequencing methods, one of ordinary skill in the art is able to identify, clone and sequence all "representative fragments" of interest including transcripts/exons encoding a large variety of human proteins. In very rare instances, this may reveal a nucleotide sequence error present in the nucleotide sequence disclosed herein. Thus, once the present invention is made available (i.e., the information in the Sequence Listing and Table 1 in a useable form), resolving a rare sequencing error would be well within the skill of the art. Nucleotide sequence editing software is publicly available.

Even if all of the very rare sequencing errors in the sequences herein disclosed were corrected, the resulting nucleotide sequence would still be at least 90% identical, and more likely 99% identical, and most likely 99.99% identical to the nucleotide sequence provided herein.

Thus, the present invention further provides nucleotide sequences that are at least 90% identical, or greater, to the nucleotide sequences of the present invention in a form that can be readily used, analyzed and interpreted by a skilled artisan. Methods for determining whether a nucleotide sequence is at least 90% identical to the nucleotide sequence of the present invention are routine and readily available to a skilled artisan. For example, the well known BLAST algorithm can be used to generate the percent identity of nucleotide sequences.

The present invention also encompasses novel amino acid sequences/proteins/peptides encoded by the transcripts/exons provided herein. Although these encoded amino acid sequences are not explicitly given, such amino acid sequences can readily be determined using the transcript/exon sequences and structural information provided herein in combination with the universal genetic code. Amino acid sequences can be readily generated by numerous algorithms or computer programs commonly used in the art that simply translate the protein-coding nucleic acid sequences provided herein into amino acid sequences based on the universal genetic code. Such amino acid/peptide sequences have commercially valuable uses similar to those described herein for the transcript/exon nucleic acid sequences/fragments of the present invention, such as design of protein detection reagents and computer-based biological analysis, for identification of commercially important proteins.

Nucleic Acid Fragments

Another embodiment of the present invention is directed to isolated fragments of the human genome, particularly those in the form of detection elements or sets of detection elements. The fragments of the human genome of the present invention include, but are not limited to, fragments that encode peptides, particularly genes, exons, and transcripts identified and described in the Sequence Listing (file SEQLIST.TXT) and in Table 1 (file TABLE1.TXT), provided on the accompanying CD labeled CL001101CDA. Such isolated fragments of the human genome, comprising the exon and/or transcript sequences provided herein and fragments thereof, are particularly useful as detection elements, such as for use as probes in a nucleic acid array, for detecting gene expression and other uses.

For example, the nucleic acid molecules/fragments of the present invention, corresponding to the transcript/exon sequences provided herein, are useful as probes, primers, chemical intermediates, and in biological assays for genes of the present invention, particularly

gene expression assays. The probes/primers can correspond to one or more of the exons provided in Table 1, or one or more of the transcripts provided in the Sequence Listing, or may span one or more exon-exon junctions identified in Table 1, or can correspond to a specific region 5' and/or 3' to a transcript or exon provided herein. The transcript/exon sequences and structural information provided herein are also useful for isolating or amplifying any given exon or transcript/gene fragment of the present invention and for designing a variety of gene, or gene expression, detection reagent/kits.

A probe/primer may comprise, for example, a substantially purified exon or transcript molecule or an oligonucleotide or oligonucleotide pair that flanks a defined transcript/exon sequence. A probe/primer comprising an exon or transcript molecule may comprise the full-length exon or transcript sequence, as provided herein, or any portion thereof. A probe/primer comprising an exon or transcript sequence may also include 5' or 3' flanking nucleic acid sequences, depending on the particular assay. Oligonucleotide probes/primers may be shorter molecules that comprise a nucleotide sequence that hybridizes under stringent conditions to at least about 5, 12, 20, 25, 40, 50, 100 or more consecutive nucleotides that comprise a unique sequence specific to the target exon or transcript/gene. Depending on the particular application, the consecutive nucleotides can either include the target exon or transcript, or be a specific region in close enough proximity 5' and/or 3' to the exon or transcript to carry out the desired assay.

Furthermore, a preferred class of nucleic acid fragments are those that span exon-exon junctions. Preferred fragments span one exon-exon junction. However, fragments may span any number of exon-exon junctions within a transcript. Nucleic acid fragments that span exon-exon junctions are particularly useful, when used as detection elements such as probes in an array, in that they eliminate false signals caused by genomic contamination. Exon-exon junctions are identified in Table 1. Nucleic acid fragments spanning exon-exon junctions can readily be determined using the exon coordinates provided in Table 1 along with the transcript sequences provided in the Sequence Listing.

The isolated nucleic acid molecules of the present invention include, but are not limited to, double-stranded or single-stranded DNA or RNA, such as mRNA, cDNA, or genomic DNA comprising the exons or transcript sequences provided herein. Isolated nucleic acid molecules may be obtained, for example, by cloning or PCR amplification, or produced by chemical synthetic techniques or by a combination thereof. Single-stranded nucleic acid can be the coding strand (sense strand) or the non-coding strand (anti-sense strand). Double-stranded RNA molecules are useful for, for example, RNA interference, or gene silencing, which can be used to

turn genes off in order to elucidate their function and may be useful therapeutic agents for turning off defective, disease-causing genes (see Plasterk *et al., Curr Opin Genet Dev* 2000 Oct;10(5):562-7; Bosher *et al., Nat Cell Biol* 2000 Feb;2(2):E31-6; and Hunter, *Curr Biol* 1999 Jun 17;9(12):R440-2).

"Nucleotide sequence" may refer to either a heteropolymer of deoxyribonucleotides, in the case of DNA, or a heteropolymer of ribonucleotides, in the case of RNA. DNA or RNA segments may be assembled, for example, from fragments of the human genome or single nucleotides, short oligonucleotide linkers, or from a series of oligonucleotides, to provide a synthetic nucleic acid molecule.

The present invention provides isolated nucleic acid molecules that contain one or more exons or transcripts disclosed by the present invention. Such nucleic acid molecules will consist of, consist essentially of, or comprise one or more exons or transcripts of the present invention. The nucleic acid molecule can have additional nucleic acid residues, such as nucleic acid residues that are naturally associated with it or heterologous nucleotide sequences.

As used herein, an "isolated" nucleic acid molecule is one that contains an exon and/or transcript of the present invention and is separated from other nucleic acid present in the natural source of the nucleic acid. The isolated nucleic acid, as used herein, will be comprised of one or more exons and/or transcripts disclosed by the present invention. The isolated nucleic acid may have flanking nucleotide sequence on either side of the exon or transcript depending on the particular use of the isolated nucleic acid or assay involved. The flanking sequence may be, for example, up to about 5,000 bases; 2,500 bases; 1,000 bases; 500 bases; 100 bases, 50 bases, 30 bases, 20 bases, or 10 bases on either side of an exon or transcript, for detection reagents. The important point is that the nucleic acid is isolated from remote and unimportant flanking sequences and is of appropriate length such that it can be subjected to the specific manipulations or uses such as recombinant expression, preparation of probes and primers for expression analysis, and other uses specific to the transcript/exon sequences.

As used herein, an "isolated nucleic acid molecule" or an "isolated fragment of the human genome" refers to a nucleic acid molecule possessing a specific nucleotide sequence which has been subjected to purification means to reduce, from the composition, the number of compounds which are normally associated with the composition. A variety of purification means that are well known in the art can be used to generate the isolated fragments of the present invention. These include, but are not limited to, methods that separate constituents of a solution based on charge, solubility, or size. Moreover, an "isolated" nucleic acid molecule, such as an mRNA molecule containing a transcript sequence of the present invention or an exon isolated from

genomic DNA, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or chemical precursors or other chemicals when chemically synthesized. However, the nucleic acid molecule can be fused to other coding or regulatory sequences and still be considered isolated. For example, recombinant DNA molecules contained in a vector are considered isolated. Further examples of isolated DNA molecules include recombinant DNA molecules maintained in heterologous host cells or purified (partially or substantially) DNA molecules in solution. Isolated RNA molecules include in vivo or in vitro RNA transcripts comprising the sequences of the present invention. Isolated nucleic acid molecules according to the present invention further include such molecules produced synthetically.

In one embodiment, human DNA can be mechanically sheared to produce fragments of about 2kb, 10kb, or 15-20 kb in length. These fragments can then be used to generate a human library by inserting them into plasmid vectors (or lambda vectors) using methods well known in the art. Primers flanking, for example a gene or exon, can then be generated using nucleotide sequence information provided in the present invention. PCR cloning can then be used to isolate the gene or exon from the human DNA library. PCR cloning is well known in the art. Thus, given the availability of the present identified gene coding sequences of the human genome, it is routine experimentation to isolate any gene or exon, or fragments thereof, particularly using the information provided in file, TABLE1.TXT, provided on the accompanying CD labeled CL001101CDA. Particularly useful is the generation of nucleic acid fragments comprising one or more exons of a gene, particularly those identified herein. Such fragments can be applied to an array, microfluidic device or other detection kit format and used to detect expression of a gene (see below).

The sequences falling within the scope of the present invention are not limited to the specific sequences herein described, but also include allelic and species variations thereof. Allelic and species variations can be routinely determined by comparing the sequences provided by the present invention, or a representative fragment thereof, with sequences from other isolates from the same species (allelic variations) or from other species (species variations). Sequence comparisons with other nucleic acid isolates to determine allelic or species variation can be readily accomplished using the transcript/exon sequences and structural information provided herein. For example, primers for re-sequencing any particular transcript, exon, or fragment thereof can be readily designed based on the sequences provided herein. Such re-sequencing is useful for detecting polymorphisms, such as SNPs, in the transcripts/exons provided herein. Furthermore, such SNPs, being in protein coding regions, are of significant commercial value

since they may change the encoded protein sequence and thereby play a direct role in disease development and progression. Such SNPs are important targets for therapeutic/drug development, and may also serve as important diagnostic/prognostic markers. Thus, the transcript/exon sequences and structural information provided herein is a commercially valuable resource for SNP detection.

To accommodate codon variability, the present invention also encompasses nucleic acid molecules coding for the same amino acid sequences as do the specific transcript/exon sequences disclosed herein. In other words, in the transcript/exon sequences disclosed herein, substitution of one codon for another that encodes the same amino acid is expressly contemplated.

The present invention further provides related nucleic acid molecules that hybridize under stringent conditions to the nucleic acid molecules disclosed herein. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences encoding a peptide at least 60-70% homologous to each other typically remain hybridized to each other. The conditions can be such that sequences at least about 60%, at least about 70%, or at least about 80%, or at least about 90% or more homologous to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in Current Protocols in Molecular Biology, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. One example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50-65°C. Examples of moderate to low stringency hybridization conditions are well known in the art.

Any specific sequence disclosed herein can be readily screened for errors by resequencing a particular fragment, such as an exon or transcript, in both directions (i.e., sequence both strands). Alternatively, error screening can be performed by sequencing corresponding polynucleotides of human origin isolated by using part or all of the fragments in question as a probe or primer.

Each of the transcripts/exons of the human genome, including sequences and isolated nucleic acid molecules, can be routinely characterized using the computer system of the present invention and can be used in numerous ways as polynucleotide reagents. For example, isolated nucleic acid molecules comprising at least one of the exon or transcript sequences provided herein, can be used as diagnostic probes or diagnostic amplification primers to detect the expression of a particular exon, exon set, gene, or gene set. This is particularly useful in the form of nucleic acid arrays wherein 100 or more, 1000 or more, 5000 or more, 10,000 or more,

or even most to all of the exons/transcripts provided by the present invention are implemented in a single array.

Nucleic Acid Arrays and Detection Kits and Reagents

The present invention provides detection kits and reagents, such as, but not limited to, arrays, TaqMan probe/primer sets, and various compartmentalized kits, comprising detection elements, such as nucleic acid probes, that are based on the sequence information provided by the present invention, particularly the transcript sequences (SEQ ID NOS:1-39010) or exon sequences (exon information is provided in Table 1).

As used herein "Arrays" or "Microarrays" refers to an array of distinct polynucleotides or oligonucleotides synthesized on a substrate, such as paper, nylon or other type of membrane, filter, chip, glass slide, plastic, silicon, gold, gel or any other suitable solid, or semi-solid support. Arrays may also be based on fiber-optics and comprise, for example, probes attached to beads at the ends of fiber-optic bundles (see Walt, *Science* 287, 451 (2000), Michael *et al., Anal. Chem* 70, 1242-1248 (1998), and Ferguson *et al., Nature Biotechnology* 14, 1681-1684 (1996)). In one embodiment, the microarray is prepared and used according to the methods described in US Patent 5,837,832 (Chee et al.), PCT application WO95/11995 (Chee et al.), Lockhart, D. J. et al. (1996; *Nat. Biotech.* 14: 1675-1680) and Schena, M. et al. (1996; *Proc. Natl. Acad. Sci.* 93: 10614-10619), all of which are incorporated herein in their entirety by reference. In other embodiments, such arrays are produced by the methods described by Brown et. al., US Patent No. 5,807,522. Hybridization and scanning of arrays is also described in PCT application WO 92/10092 and EP785280. The use of microarrays of oligonucleotides or polynucleotides for capturing complementary polynucleotides from expressed genes is also described in Schena *et al, Science*, 270: 467-469 (1995); DeRisi *et al, Science*, 278: 680-686 (1997); Chee *et al, Science*, 274: 610-614 (1996). Additionally, Freeman *et al.* (*Biotechniques* 29, 1042-1055 (2000), Lockhart *et al.* (*Nature* 405, 827-836 (2000)), and Zweiger (*Trends in Biotechnology* 17, 429-436 (1999)) provide reviews of nucleic acid arrays for gene expression analysis and other uses; also see *Nature Genetics* 21 (Suppl.), 1-60 (1999) and Meldrum, *Genome Research*, 10:1288-1303 (2000) for an overview of array technology.

For example, gene expression kits and reagents, such as arrays or sets of probe containing beads, may contain one or more detection elements, such as oligonucleotide probes or pairs of probes, that hybridize at or near each exon or gene corresponding to the exon/transcript sequences provided by the present invention. A plurality of oligonucleotide probes may be included in the kit to simultaneously assay large numbers of genes/exons, at least one of which is

one of the genes/exons of the present invention and novel to the present disclosure. In some kits, such as arrays, the oligonucleotide probes are provided immobilized to a substrate. For example, the same substrate can comprise oligonucleotide probes for detecting at least 1; 10; 100; 1000; 10,000 or most or substantially all of the genes/transcripts or exons provided by the present invention. Any number of probes, or other detection elements, may be utilized in a detection reagent, depending on the particular technology platform and objective. For example, a typical array may contain hundreds, thousands to millions of individual synthetic DNA probes arranged in a grid-like pattern and miniaturized to the size of a dime, each corresponding to a particular exon or transcript/gene. Preferably, probes are attached to a solid support in an ordered, addressable array. Customized arrays that utilize the exon and/or gene/transcript sequences provided by the present invention can be produced by various manufacturers. For example, arrays with over 250,000 oligonucleotide probes or 10,000 cDNAs per square centimeter are readily available (see Lipshutz *et al., Nature Genetics*, 21, 20-24 (1999) and Bowtell *et al., Nature Genetics*, 21, 25-32 (1999). In some arrays, electric fields can be applied to the array to speed hybridization reactions (see Edman *et al., Nucleic Acids Res.* 25, 4907-4914 (1997) and Sosnowski *et al., Proc. Natl. Acad. Sci. USA* 94, 1119-1123 (1997)). Arrays have been previously produced for completely sequenced organisms, such as *Saccharomyces cerevisiae*, that comprise probes for every identified gene in the organism's genome (see DeRisi *et al., Science* 278, 680-686 (1997) and Wodicka *et al., Nature Biotechnology* 15, 1359-1367 (1997)).

The microarray or detection kit is preferably composed of a large number of unique nucleic acid sequences, usually either synthetic antisense oligonucleotides or fragments of cDNAs, fixed to a solid support. Probes may comprise either single- or double-stranded nucleic acid molecules. Oligonucleotides may be about 6-60 nucleotides in length, more preferably 15-30 nucleotides in length, and most preferably about 20-25 nucleotides in length. For a certain type of microarray or detection kit, it may be preferable to use oligonucleotides that are only 7-20 nucleotides in length. For others, such as cDNA, longer lengths are possible and preferable. These can be of the order of 1kb-5kb or more in length and can comprise the entire length of a transcript or exon sequence provided herein or can comprise a short fragment of the transcript/exon, such as in exon-exon junction spanning detection elements.

The microarray or detection kit may contain oligonucleotides that cover, for example, sequential oligonucleotides that cover the full-length sequence, or unique oligonucleotides selected from particular areas along the length of the sequence, such as in exon-exon boundaries. Additionally, such as in the case of primers for PCR, it may be desirable for oligonucleotides to bind to regions 5' or 3' of the transcripts/exons provided herein, such as to capture the entire

exon or transcript/gene within the amplicon. Polynucleotides used in the microarray or detection kit may be oligonucleotides that are specific to an exon, exons, gene, or genes of interest.

Thus, the chip may comprise an array comprising at least one probe corresponding to the full-length sequence of at least one of the exons and/or transcripts provided by the present invention, sequences spanning one or more exon-exon junctions identified in Table 1, sequences complementary thereto, or fragments thereof. Thus, the sequence of at least one probe of the array is selected from the group consisting of those disclosed in SEQ ID NOS:1-39010 and the exons identified in Table 1, sequences spanning one or more exon-exon junctions identified in Table 1, sequences complementary thereto, and fragments thereof.

In order to produce oligonucleotides to a known sequence for a microarray or detection kit, the exon(s) or gene(s) of interest is typically examined using a computer algorithm that starts at the 5' or at the 3' end of the nucleotide sequence. Typical algorithms will then identify oligomers of defined length that are unique to the exon/gene, have a GC content within a range suitable for hybridization, and lack predicted secondary structure that may interfere with hybridization. In certain situations it may be appropriate to use pairs of oligonucleotides on a microarray or detection kit. For example, pairs of oligonucleotides are particularly useful for detecting mismatch hybridization in high-density arrays that use short oligonucleotides, such as 25-mers; such short oligonucleotides are susceptible to mismatch hybridization due to false priming. In this situation, pairs of oligonucleotides with deliberate mismatches are incorporated to determine the level of mismatch hybridization, which can then be subtracted from the true target signal (Lockhart *et al., Nat. Biotechnology* (1996) 14:1675-1680 and Wodicka *et al., Nat. Biotechnology* (1997) 15:1359-1366). Pairs of oligonucleotide probes are also useful for detecting polymorphisms, particularly SNPs; in these situations, the oligonucleotide pairs are generally designed to be identical except for one nucleotide that preferably is located at or near the center of the sequence. The second oligonucleotide in the pair (mismatched by one) serves as a control. The oligomers are synthesized at designated areas on a substrate using a light-directed chemical process. The substrate may be paper, nylon or other type of membrane, filter, chip, glass slide or any other suitable solid support.

In another aspect, an oligonucleotide may be synthesized on the surface of the substrate by using a chemical coupling procedure and an ink jet application apparatus, as described in PCT application WO95/251116 (Baldeschweiler et al.) which is incorporated herein in its entirety by reference. In another aspect, a "gridded" array analogous to a dot (or slot) blot may be used to arrange and link cDNA fragments or oligonucleotides to the surface of a substrate using, for example, a vacuum system, thermal, UV, mechanical or chemical bonding procedure. An array,

such as those described above, may be produced by hand or by using available devices (slot blot or dot blot apparatus), materials (any suitable solid support), and machines (including robotic instruments), and may contain 8; 24; 96; 384; 1536; 6144; 10,000 or more oligonucleotides, or any other number which lends itself to the efficient use of commercially available instrumentation.

In other embodiments, the array or detection reagent/kit can be produced by spotting cDNA or other nucleic acid molecules onto the surface of a substrate (see Brown et. al., US Patent No. 5,807,522). In such use, PCR amplification of one or more exons or transcripts from genomic DNA can be used to generate a nucleic acid molecule suitable for deposition onto a substrate.

In yet another embodiment, the detection reagent or kit comprises TaqMan probe/primer sets for carrying out the TaqMan PCR assay, such as for detecting gene expression. The TaqMan assay, also known as the 5' nuclease PCR assay, provides a sensitive and rapid means of detecting gene expression. The TaqMan assay detects the accumulation of a specific amplified product during PCR. The TaqMan assay utilizes an oligonucleotide probe labeled with a fluorescent reporter dye at the 5' end of the probe and a quencher dye at the 3' end of the probe. During the PCR reaction, the 5' nuclease activity of DNA polymerase cleaves the probe, thereby separating the reporter dye and the quencher dye and resulting in increased fluorescence of the reporter. Accumulation of PCR product is detected directly by monitoring the increase in fluorescence of the reporter dye. The 5' nuclease activity of DNA polymerase cleaves the probe between the reporter and the quencher only if the probe hybridizes to the target and is amplified during PCR. The probe is designed to hybridize to a target nucleic acid molecule only if the target sequence is complementary to the probe, i.e., if the target sequence comprises the transcript/exon sequence that is used as a probe.

Preferred TaqMan primer and probe sequences can readily be determined using the nucleic acid information provided herein. A number of computer programs, such as Primer-Express, can be used to readily obtain optimal primer/probe sets. It will be apparent to one of skill in the art that the primers and probes based on the nucleic acid and transcript/exon sequences and structural information provided herein are useful as probes or amplification primers for screening for the transcripts/exons provided by the present invention, such as for monitoring gene expression in particular disease conditions, and can be incorporated into a kit format. In particular, genome-wide TaqMan probe/primer sets are specifically contemplated for monitoring the expression of 10,000 or more, or most or all, human genes, or any subset thereof of interest. Such genome-wide TaqMan probe/primer sets can readily be obtained using the

transcript sequences and transcript/exon structural information provided herein, along with a primer/probe design computer program, such as Primer-Express.

Other detection kits and reagents may be based on blotting techniques such as northern blots (for detecting RNA), southern blots (for detecting DNA), or western blots (for detecting proteins) or beads containing detection elements that are well known in the art. The exons and transcript sequences provided by the present invention are well suited for use as detection probes in such techniques.

Direct sequencing, including cDNA sequencing, can also be used to detect the transcripts and/or exons of the present invention. A variety of automated sequencing procedures can be utilized when performing detection/diagnostic assays ((1995) *Biotechniques* 19:448), including sequencing by mass spectrometry (see, e.g., PCT International Publication No. WO 94/16101; Cohen *et al.*, *Adv. Chromatogr.* 36:127-162 (1996); and Griffin *et al.*, *Appl. Biochem. Biotechnol.* 38:147-159 (1993)).

Various other methods useful for gene expression analysis include, but are not limited to, RT-PCR, nuclease protection, clone hybridization, differential display (Liang *et al.*, *Science* 257, 967-971 (1992)), subtractive hybridization, cDNA fingerprinting (Shimkets *et al.*, *Nature Biotechnology* 17, 798-803 (1999), Ivanova, *Nucleic Acids Research* 23, 2954-2958 (1995), Kato, *Nucleic Acids Research* 23, 3685-3690 (1995), and Bachem *et al.*, *Plant J.* 9, 745-753 (1996)), reporter-gene analysis, two-dimensional (2D) gel electrophoresis, mass spectrometry, and serial analysis of gene expression (SAGE) (Velculescu *et al.*, *Science* 270, 484-487 (1995)).

In order to conduct sample analysis using a microarray or other detection reagent/kit, a typical procedure may be similar to the following. The RNA or DNA from a biological sample is made into hybridization probes. The mRNA is isolated, and cDNA is produced and used as a template to make antisense RNA (aRNA). The aRNA is amplified in the presence of fluorescent nucleotides, and labeled probes are incubated with the microarray or detection kit so that the probe sequences hybridize to complementary oligonucleotides of the microarray or detection kit. Incubation conditions may be adjusted so that hybridization occurs with precise complementary matches or with various degrees of less complementarity. After removal of nonhybridized probes, a scanner is used to determine the levels and patterns of fluorescence. The scanned images are examined to determine degree of complementarity and the relative abundance of each oligonucleotide sequence on the microarray or detection kit. The biological samples may be obtained from any bodily fluids (such as blood, urine, saliva, phlegm, gastric juices, etc.), cultured cells, biopsies, or other tissue preparations. A detection system may be used to measure the absence, presence, and amount of hybridization for all of the distinct sequences

simultaneously. This data may be used for purposes including, but not limited to, large-scale correlation studies on the sequences, expression patterns, mutations, variants, or polymorphisms among samples.

Using such arrays, the present invention provides methods to identify the expression of one or more of the exons or transcripts/genes of the present invention. Such methods may comprise incubating a test sample with an array comprising one or more oligonucleotide probes corresponding to at least one exon or transcript of the present invention and assaying for binding of a nucleic acid from the test sample with one or more of the oligonucleotide probes. Such assays will typically involve arrays comprising most, if not all, of the exons or transcripts in the human genome, or rationally selected subsets thereof. The transcript sequences of the human genome are provided in SEQ ID NOS:1-39010 and the exons that these transcripts are comprised of are provided in Table 1.

Conditions for incubating a nucleic acid molecule with a test sample vary. Incubation conditions depend on the format employed in the assay, the detection methods employed, and the type and nature of the nucleic acid molecule used in the assay. One skilled in the art will recognize that any one of the commonly available hybridization, amplification, or array assay formats can readily be adapted to employ the novel fragments of the human genome disclosed herein. Examples of such assays can be found in Chard, T, *An Introduction to Radioimmunoassay and Related Techniques,* Elsevier Science Publishers, Amsterdam, The Netherlands (1986); Bullock, G. R. *et al., Techniques in Immunocytochemistry,* Academic Press, Orlando, FL Vol. 1 (1982), Vol. 2 (1983), Vol. 3 (1985); Tijssen, P., *Practice and Theory of Enzyme Immunoassays: Laboratory Techniques in Biochemistry and Molecular Biology,* Elsevier Science Publishers, Amsterdam, The Netherlands (1985).

The test samples of the present invention include, but are not limited to, nucleic acid extracts, cells, and protein or membrane extracts from cells, which may be obtained from any bodily fluids (such as blood, urine, saliva, phlegm, gastric juices, etc.), cultured cells, biopsies, or other tissue preparations. The test sample used in the above-described methods will vary based on the assay format, nature of the detection method and the tissues, cells or extracts used as the sample to be assayed. Methods of preparing nucleic acid, protein, or cell extracts are well known in the art and can be readily be adapted in order to obtain a sample that is compatible with the system utilized.

In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out one or more assays for detecting the exons/transcripts/genes of the present invention, such as for gene expression analysis. Specifically, the invention provides a

compartmentalized kit to receive, in close confinement, one or more containers, comprising: (a) a first container comprising at least one nucleic acid molecule that can bind to a fragment of at least one of the exon or transcript sequences disclosed herein, including exon-exon spanning sequences; and (b) one or more other containers comprising wash reagents and/or reagents capable of detecting presence of a bound nucleic acid. Preferred kits will include detection reagents/arrays/chips/microfluidic devices that are capable of detecting the expression of 1 or more, 10 or more, 100 or more, 500 or more, 1000 or more, 10,000 or more, or most or all of the exons or transcripts identified herein that are expressed in humans. One skilled in the art will readily recognize that the previously unidentified exons/transcripts provided by the present invention can be readily incorporated into one of the established kit formats which are well known in the art, particularly expression arrays.

In detail, a compartmentalized kit includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers, strips of plastic, glass or paper, or arraying material such as silica. Such containers allow one to efficiently transfer reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated, and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another. Such kits may typically include a container which will accept the test sample, a container which contains the nucleic acid probe, containers which contain wash reagents (such as phosphate buffered saline, Tris-buffers, etc.), and containers which contain the reagents used to detect the bound probe. The kit can further comprise reagents for PCR, RT-PCR or other enzymatic reactions, and instructions for using the kit. Such compartmentalized kits include multicomponent integrated systems.

Multicomponent integrated systems may also implement the transcript/exon sequences, including exon-exon spanning sequences, provided by the present invention as detection elements. Multicomponent integrated systems include such systems as microfluidic devices, biomedical micro-electro-mechanical systems (bioMEMS), and "lab-on-a-chip" systems (see, for example, US patents 6,153,073, Dubrow *et al.*, and 6,156,181, Parce *et al.*). Such systems miniaturize and compartmentalize processes such as probe/target hybridization, PCR, and capillary electrophoresis reactions in a single functional device, and may be integrated with nucleic acid arrays. An example of such a technique is disclosed in US patent 5,589,136, which describes the integration of PCR amplification and capillary electrophoresis in chips. Multicomponent integrated systems such as microfluidic, bioMEMs, and lab-on-a-chip systems, generally comprise a pattern of microchannels designed onto a glass, silicon, quartz, or plastic wafer included on a microchip. The movements of the samples are controlled by electric,

electroosmotic, or hydrostatic forces applied across different areas of the microchip to create functional microscopic valves and pumps with no moving parts. Varying the voltage controls the liquid flow at intersections between the micro-machined channels and changes the liquid flow rate for pumping across different sections of the microchip.

Medical- and Pharmaceutical-Related Uses

Detection of gene expression, using the transcripts and/or exons of the present invention, is valuable for such uses as disease diagnosis, monitoring disease progression, determining the effects of various treatments/therapeutics, and individualizing medical treatment or drug therapy based on an individual's gene expression patterns. In particular, uses such as these can be achieved using the detection reagents provided by the present invention, such as nucleic acid arrays that utilize the human exon and/or transcript sequences provided by the present invention as detection elements. Genome-wide expression analysis can be conducted in humans using the exons/transcripts provided herein; genome-wide expression analysis has previously been accomplished in yeast (Holstege *et al., Cell* 95, 717-728 (1998)).

Detection reagents, such as arrays, containing the transcripts/exons of the present invention can also be used to probe genomic DNA for changes in gene copy number or allelic imbalances (see Mei *et al., Genome Res* 2000 Aug;10(8):1126-37, Pollack *et al., Nature Genetics* 23, 41-46 (1999), and Pinkel *et al. Nature Genetics* 20, 207-211 (1998)). Such copy number changes/allelic imbalances may be caused by gene or chromosome deletions or duplications, which may occur in cancerous cells and other disorders. Furthermore, identification of genetic/chromosomal changes such as these may facilitate the identification of specific genes, regulatory/control regions, or other genetic elements that play important roles in the disorder, or indicate that a particular chromosomal region harbors such elements.

The sequences and detection reagents of the present invention may be used to determine whether an individual has a mutation or polymorphism, such as a SNP (single nucleotide polymorphism), affecting the level (i.e., the concentration of mRNA or protein in a sample, etc.) or pattern (i.e., the kinetics of expression, rate of decomposition, stability profile, Km, Vmax, etc.) of gene expression in a particular cell, tissue, bodily fluid, disease state, or developmental stage. Such variations in gene expression can be caused, for example, by a SNP in a gene, or in a regulatory/control region(s), such as a promoter, or other gene(s) that controls or affects the expression of the gene. Such an analysis of gene expression can be conducted by screening for mRNA corresponding to the exons and/or transcripts provided by the present invention. Once changes in gene expression patterns are identified, the nucleic acid sequences provided by the

present invention can be used, for example, to design primers/probes for SNP-detection assays to determine if a SNP is responsible for the variation in gene expression patterns. Such SNP-detection assays include, but are not limited to, direct sequencing, mini-sequencing primer extension, and the TaqMan PCR assay, or any other SNP-detection technique known in the art. Furthermore, SNP-detection assays may utilize nucleic acid arrays, mass spec, or other technology platforms used in the art for SNP-detection. Once a SNP is detected that alters gene expression in a manner that contributes to a pathological condition, therapeutic approaches can be targeted at that SNP and, furthermore, that SNP can serve as a diagnostic/prognostic marker for the disease, and may form the basis of a diagnostic kit for the disease. Furthermore, SNPs in the transcript/exon coding sequences provided herein can readily be determined by comparing the sequences provided herein against corresponding transcript/exon sequences from nucleic acid isolates taken from different individuals, such as by re-sequencing or computer-based sequence database comparison. Additionally, changes in the amino acid/protein sequences caused by such SNPs can readily be determined using the sequences provided herein as a reference and the universal genetic code.

Medical gene expression analysis can include the steps of collecting a sample of cells from a patient, isolating mRNA from the cells of the sample, contacting the mRNA sample with one or more probes, based on the exon and/or transcript sequences provided herein, which specifically hybridize to a region of the isolated mRNA containing a target exon/transcript under conditions such that hybridization of the probe with the exon/transcript occurs, and detecting the presence or absence of hybridization. The presence or absence of hybridization, and therefore of the target exon/transcript, can then be correlated with known gene expression patterns in, for example, normal cells/tissues and in cells/tissues in various disease stages in order to, for example, diagnose a disease, determine disease progression, or determine the effect of a particular drug treatment.

The contribution or association of particular gene expression patterns with disease phenotypes enables the transcripts/exons of the present invention to be used to develop superior diagnostic tests based on gene expression/mRNA markers. Such gene expression-based diagnostic tests are useful for identifying individuals who have a gene expression indicative of a specific disease or disease propensity or individuals whose gene expression patterns indicate that a particular drug treatment or therapeutic approach should be utilized. For example, HER2 and the estrogen receptor genes are known to be expressed at increased levels in cancers, such as breast and ovarian cancer (van de Vijver *et al.* (1988) *New Engl. J. Med.* 319, 1239-1245, Berger *et al.* (1988) *Cancer Res.* 48, 1238-1243, and Petrangeli *et al.* (1994) *J. Steroid Biochem. Mol.*

Biol. 49, 327-331) and determining the expression level of these genes may aid physicians in choosing the most effective treatment (McNeil *et al.* (1999) *J. Natl. Cancer Inst.* 91, 110-112, Leinster *et al.* (1998) *Biochem Soc. Symp.* 63, 185-191, and Revillon *et al.* (1998) *Eur. J. Cancer* 34, 791-808). Such diagnostics may be based on a single transcript/gene or exon, a group of transcripts/genes or exons, or most or all transcripts/genes or exons provided by the present invention.

The invention further provides a method for identifying a compound that can be used to treat a disorder associated with expression of a disease-associated gene or variable, disease-associated, expression of a normal gene. Forms of gene expression such as these are collectively referred to herein as disease-associated gene expression, and may contribute to, for example, disease or developmental disorders. The method typically includes assaying the ability of the compound to modulate the activity and/or expression of the target nucleic acid and thus identifying a compound that can be used to treat a disorder characterized by undesired activity or expression of the nucleic acid.

The assays for disease-associated nucleic acid expression can be accomplished using the transcript and/or exon sequences provided by the present invention as gene expression detection elements, such as probes in a nucleic acid array. The assay for disease-associated nucleic acid expression can involve direct assay of nucleic acid levels, such as mRNA levels, or on collateral compounds involved in the signal pathway. Further, the expression of genes that are up- or down regulated in response to the disease-associated protein signal pathway can also be assayed. In this embodiment the regulatory regions of these genes can be operably linked to a reporter gene such as luciferase.

Thus, modulators of disease-associated gene expression can be identified in a method wherein a cell is contacted with a candidate compound, such as a drug or small molecule, and the expression of mRNA determined. The level of expression of disease-associated mRNA in the presence of the candidate compound is compared to the level of expression of disease-associated mRNA in the absence of the candidate compound. The candidate compound can then be identified as a modulator of nucleic acid expression based on this comparison and be used, for example, to treat a disorder characterized by disease-associated gene expression. When expression of mRNA is statistically significantly greater in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of nucleic acid expression. When nucleic acid expression is statistically significantly less in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of nucleic acid expression.

The invention further provides methods of treatment, with one or more of the genes/transcripts/exons provided by the present invention as a target, using a compound identified through drug screening using the transcript/exon sequences provided herein, as a gene modulator to modulate nucleic acid expression. Modulation includes both up-regulation (i.e. activation or agonization) or down-regulation (suppression or antagonization) of nucleic acid expression. These methods of treatment include the step of administering the modulators of gene expression in a pharmaceutical composition to a subject in need of such treatment.

The exon/transcript sequences provided herein are also useful for monitoring the effectiveness of modulating compounds on the expression or activity of a gene in clinical trials or in a treatment regimen. Thus, the gene expression pattern can serve as a barometer for the continuing effectiveness of treatment with the compound, particularly with compounds to which a patient can develop resistance. The gene expression pattern can also serve as a marker indicative of a physiological response of the affected cells to the compound. Accordingly, such monitoring would allow either increased administration of the compound or the administration of alternative compounds to which the patient has not become resistant. Similarly, if the level of nucleic acid expression falls below a desirable level, administration of the compound could be commensurately decreased. Therefore, the transcript/exon sequences of the present invention are particularly useful for improving the process of drug development by allowing changes in gene expression patterns in response to candidate compounds/drugs to be determined; such changes in gene expression patterns can be analyzed to determine compound/drug efficacy and/or toxicity. This not only improves the safety of clinical trials, but also will enhance the chances that the trial will demonstrate statistically significant efficacy by allowing the clinical trials to be adjusted in response to different gene expression patterns observed in different patients in response to a candidate compound/drug. Furthermore, gene expression analysis using the transcripts/exons of the present invention may help explain why certain, previously developed drugs performed poorly in clinical trials and may help identify a subset of the population that would benefit from a drug that had previously performed poorly in clinical trials, thereby "rescuing" previously developed drugs.

Gene expression analysis using the detection reagents of the present invention is also useful for determining the target of a drug. For example, gene expression patterns in cells treated with a drug can be compared to gene expression patterns in cells that have had individual genes, particularly genes corresponding to the exons/transcripts provided herein, inactivated. A similar gene expression pattern would indicate that the drug may target the gene that had been inactivated.

Gene expression analysis, using the transcripts/exons provided by the present invention, may also be useful in forensic and medicolegal investigations. For example, post-mortem gene expression analysis may provide clues as to cause of death or time of death, may indicate exposure to toxic compounds or drugs, and may aid in identification.

Examples of other important uses of the transcripts/exons provided herein for gene expression include, but are not limited to, determining the toxicological consequences of altered gene expression (Pennie, *Toxicol Lett* 2000 Mar 15; 112-113: 473-7), understanding changes in gene expression in response to infection (Manger *et al., Curr Opin Immunol* 2000 Apr;12(2):215-8) and modulating gene expression to enhance the immune response, and regulating the expression of genes delivered through gene therapy (Clackson, *Gene Ther* 2000 Jan;7(2):120-5).

Expression Modulating Fragments

The present invention is useful for unraveling and characterizing the complex genetic network involved in the regulation and control of gene expression. For example, the present invention facilitates the identification and characterization of regulatory/control elements in the human genome, referred to herein as "expression modulating fragments" (EMFs), or expression modulating elements/sequences. As used herein, an "expression modulating fragment," means a series of nucleotide molecules that modulate the expression of an operably linked gene/transcripts or another EMF. EMFs may also include gene products such as transcriptional activators and repressors. Sets of co-regulated genes, referred to as "regulons", can also be identified. Genomic features such as novel EMFs and regulons can be identified, for example, through genome-wide expression analysis using arrays comprising the exons/transcripts provided by the present invention. Genomic sequence motifs that are statistically over-abundant in regions close to similarly expressed genes, particularly in 5' regions, may be identified as novel EMFs, such as *cis*-regulatory elements. Furthermore, using genome-wide expression analysis, one can determine whether an EMF has a global effect (affects a large number of genes, or all genes) or a specific effect (affects a small number of genes, or a single gene) (Holstege *et al., Cell* 95, 717-728 (1998)). Additionally, by providing a tool for monitoring gene/transcript expression, the present invention is also useful for monitoring variations in gene/transcript expression in response to known mutations or polymorphisms in EMFs, or to identify previously unknown mutations or polymorphisms in EMFs based on variations in gene expression. Such polymorphisms in EMFs, particularly SNPs, may be useful diagnostic markers for disease.

As used herein, a sequence is said to "modulate the expression of an operably linked sequence" when the expression of the sequence is altered by the presence of the EMF. EMFs include, but are not limited to, promoters, and promoter modulating sequences (inducible elements). One class of EMFs is comprised of fragments that induce the expression of an operably linked gene/transcript in response to a specific regulatory factor or physiological event.

EMF sequences can be identified within the human genome by their proximity to the transcripts/exons provided by the present invention. An intergenic segment, or a fragment of the intergenic segment, from about 10 to 200, 10 to 500, 10 to 1kB, or 10 to 2.5kB nucleotides in length, preferably taken 5' from any one of the transcripts identified in the Sequence Listing (file SEQLIST.TXT), provided on the accompanying CD labeled CL001101CDA, will modulate the expression of an operably linked 3' gene/transcript in a fashion similar to that found with the naturally linked gene/transcript sequence. As used herein, an "intergenic segment" refers to fragments of the human genome that are between two transcripts herein described. Alternatively, EMFs can be identified using known EMFs as a target sequence or target motif in the computer-based systems of the present invention.

The presence and activity of an EMF can be confirmed using an EMF trap vector. An EMF trap vector contains a cloning site 5' to a marker sequence. A marker sequence encodes an identifiable phenotype, such as antibiotic resistance or a complementing nutrition auxotrophic factor, which can be identified or assayed when the EMF trap vector is placed within an appropriate host under appropriate conditions. As described above, an EMF will modulate the expression of an operably linked marker sequence. A sequence which is suspected as being an EMF is cloned in all three reading frames in one or more restriction sites upstream from the marker sequence in the EMF trap vector. The vector is then transformed into an appropriate host using known procedures and the phenotype of the transformed host in examined under appropriate conditions. As described above, an EMF will modulate the expression of an operably linked marker sequence.

Computer Related Embodiments

The nucleotide sequences provided by the present invention, a representative fragment thereof, or nucleotide sequences at least 99% identical to these sequences, may be "provided" in a variety of mediums to facilitate use thereof. As used herein, "provided" refers to a manufacture, other than an isolated nucleic acid molecule, that contains a nucleotide sequence of the present invention, i.e., the nucleotide sequences provided in the present invention, a representative fragment thereof, or nucleotide sequences at least 99% identical to these

sequences. Such a manufacture provides the coding portion of the human genome or a subset thereof (e.g., a human exon or transcript sequence) in a form that allows a skilled artisan to examine the manufacture using means not directly applicable to examining the human genome or a subset thereof as it exists in nature or in purified form.

In one application of this embodiment, a nucleotide sequence of the present invention can be recorded on computer readable media. As used herein, "computer readable media" refers to any medium that can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media. A skilled artisan can readily appreciate how any of the presently known computer readable mediums can be used to create a manufacture comprising computer readable medium having recorded thereon a nucleotide sequence of the present invention. One such medium is provided with the present application, namely, the present application contains computer readable medium (CD-R) that has the transcript sequences provided/recorded thereon in ASCII text format in a Sequence Listing (provided in file SEQLIST.TXT on the accompanying CD labeled CL001101CDA.

As used herein, "recorded" refers to a process for storing information on computer readable medium. A skilled artisan can readily adopt any of the presently known methods for recording information on computer readable medium to generate manufactures comprising the nucleotide sequence information of the present invention.

A variety of data storage structures are available to a skilled artisan for creating a computer readable medium having recorded thereon a nucleotide sequence of the present invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the nucleotide sequence information of the present invention on computer readable medium. The sequence information can be represented in a word processing text file, formatted in commercially-available software such as WordPerfect and Microsoft Word, or represented in the form of an ASCII file, stored in a database application, such as OB2, Sybase, Oracle, or the like. A skilled artisan can readily adapt any number of data processor structuring formats (e.g., text file or database) in order to obtain computer readable medium having recorded thereon the nucleotide sequence information of the present invention.

By providing the nucleotide sequences of the present invention, a representative fragment thereof, or nucleotide sequences at least 99% identical to these sequences, in computer readable form, a skilled artisan can routinely access the sequence information for a variety of purposes.

Computer software is publicly available which allows a skilled artisan to access sequence information provided in a computer readable medium. Software which implements the BLAST (Altschul *et at, J. Mol. Biol.* 215:403-410 (1990)) and BLAZE (Brutlag *et at, Comp. Chem.* 17:203-207 (1993)) search algorithms on a Sybase system may be used to identify exons/transcripts within the human genome which contain homology to nucleic acid or proteins sequences from other organisms. Such exons/transcripts are protein-encoding fragments within the human genome and are useful in producing commercially important proteins such as therapeutic proteins.

The present invention further provides systems, particularly computer-based systems, which contain the sequence information described herein. Such systems are designed to identify commercially important fragments of the human genome.

As used herein, a "computer-based system" refers to the hardware means, software means, and data storage means used to analyze the nucleotide sequence information of the present invention. The minimum hardware means of the computer-based systems of the present invention comprises a central processing unit (CPU), input means, output means, and data storage means. A skilled artisan can readily appreciate that any one of the currently available computer-based systems are suitable for use in the present invention. Such system can be changed into a system of the present invention by utilizing the sequence information provided on the CD-R, or a subset thereof, without any experimentation.

As stated above, the computer-based systems of the present invention comprise a data storage means having stored therein a nucleotide sequence of the present invention and the necessary hardware means and software means for supporting and implementing a search means. As used herein, "data storage means" refers to memory which can store nucleotide sequence information of the present invention, or a memory access means which can access manufactures having recorded thereon the nucleotide sequence information of the present invention.

As used herein, "search means" refers to one or more programs that are implemented on the computer-based system to compare a target sequence or target structural motif with the sequence information stored within the data storage means. Search means are used to identify fragments or regions of the human genome that match a particular target sequence or target motif. A variety of known algorithms are disclosed publicly and a variety of commercially available software for conducting search means are available and can be used in the computer-based systems of the present invention. Examples of such software include, but are not limited to, MacPattern (EMBL), BLASTN and BLASTX (NCBIA). A skilled artisan can readily

recognize that any one of the available algorithms or implementing software packages for conducting homology searches can be adapted for use in the present computer-based systems.

As used herein, a "target sequence" can be any DNA or amino acid sequence of six or more nucleotides or two or more amino acids. A skilled artisan can readily recognize that the longer a target sequence is, the less likely a target sequence will be present as a random occurrence in the database. The most preferred sequence length of a target sequence is from about 10 to 100 amino acids or from about 20 to 300 nucleotide residues. However, it is well recognized that searches for commercially important fragments of the human genome, such as sequence fragments involved in gene expression and protein processing, may be of shorter length.

As used herein, "a target structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences in which the sequence(s) is chosen based on a three-dimensional configuration which is formed upon the folding of the target motif. There are a variety of target motifs known in the art. Protein target motifs include, but are not limited to, enzymatic active sites and signal sequences. Nucleic acid target motifs include, but are not limited to, promoter sequences, hairpin structures and inducible expression elements (protein binding sequences).

A variety of structural formats for the input and output means can be used to input and output the information in the computer-based systems of the present invention. A preferred format for an output means ranks fragments of the human genome possessing varying degrees of homology to the target sequence or target motif. Such presentation provides a skilled artisan with a ranking of sequences which contain various amounts of the target sequence or target motif and identifies the degree of homology contained in the identified fragment.

A variety of comparing means can be used to compare a target sequence or target motif with the data storage means to identify sequence fragments of the human genome. For example, software which implements the BLAST and BLAZE algorithms (Altschul *et al., J Mol. Biol.* 215:403-410 (1990)) can be used to identify sequence fragments of interest within the human genome. A skilled artisan can readily recognize that any one of the publicly available homology search programs can be used as the search means for the computer-based systems of the present invention.

One application of this embodiment is provided in the figure. The figure provides a block diagram of a computer system 102 that can be used to implement the present invention. The computer system 102 includes a processor 106 connected to a bus 104. Also connected to the bus 104 are a main memory 108 (preferably implemented as random access memory, RAM)

and a variety of secondary storage devices 110, such as a hard drive 112 and a removable medium storage device 114. The removable medium storage device 114 may represent, for example, a floppy disk drive, a CD-ROM drive, a magnetic tape drive, etc- A removable storage medium 116 (such as a floppy disk, a compact disk, a magnetic tape, etc.) containing control logic and/or data recorded therein may be inserted into the removable medium storage device 114. The computer system 102 includes appropriate software for reading the control logic and/or the data from the removable storage medium 116 once inserted in the removable medium storage device 114.

The nucleotide sequences of the present invention may be stored in a well-known manner in the main memory 108, any of the secondary storage devices 110, and/or a removable storage medium 116. Software for accessing and processing the nucleotide sequence (such as search tools, comparing tools, etc.) reside in main memory 108 during execution.

All publications and patents mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the above-described modes for carrying out the invention which are obvious to those skilled in the field of molecular biology or related fields are intended to be within the scope of the following claims.

Claims

That which is claimed is:

1) An isolated nucleic acid detection reagent that is capable of detecting the presence of 100,000 or more human exons, wherein said exons are selected from the group consisting of those identified in Table 1.
2) The detection reagent of claim 1, wherein said reagent is a nucleic acid array.
3) The array of claim 2, wherein said array is comprised of short oligonucleotides from about 5 to about 100 nucleotides in length.
4) The array of claim 2, wherein said array is comprised of polynucleotides based on the transcript sequences (SEQ ID NOS:1-39010), wherein said polynucleotides are from about 100 to about 1000 nucleotides in length.
5) An isolated nucleic acid detection reagent that is capable of detecting the presence of 2000 or more human exons, wherein said exons are selected from the group consisting of those identified in Table 1.
6) The detection reagent of claim 5, wherein said reagent is a nucleic acid array.
7) The array of claim 6, wherein said array is comprised of short oligonucleotides from about 5 to about 100 nucleotides in length.
8) The array of claim 6, wherein said array is comprised of polynucleotides based on the transcript sequences (SEQ ID NOS:1-39010), wherein said polynucleotides are from about 100 to about 1000 nucleotides in length.
9) An isolated nucleic acid detection reagent that is capable of detecting the presence of 5000 or more human exons, wherein said exons are selected from the group consisting of those identified in Table 1.
10) The detection reagent of claim 9, wherein said reagent is a nucleic acid array.
11) The array of claim 10, wherein said array is comprised of short oligonucleotides from about 5 to about 100 nucleotides in length.
12) The array of claim 10, wherein said array is comprised of polynucleotides based on the transcript sequences (SEQ ID NOS:1-39010), wherein said polynucleotides are from about 100 to about 1000 nucleotides in length.
13) An isolated nucleic acid detection reagent that is capable of detecting the presence of 10,000 or more human exons, wherein said exons are selected from the group consisting of those identified in Table 1.
14) The detection reagent of claim 13, wherein said reagent is a nucleic acid array.

15) The array of claim 14, wherein said array is comprised of short oligonucleotides from about 5 to about 100 nucleotides in length.

16) The array of claim 14, wherein said array is comprised of polynucleotides based on the transcript sequences (SEQ ID NOS:1-39010), wherein said polynucleotides are from about 100 to about 1000 nucleotides in length.

17) The detection reagent of claim 1, wherein said reagent is comprised of at least one polynucleotide spanning at least one exon-exon junction identified in Table 1.

18) The detection reagent of claim 5, wherein said reagent is comprised of at least one polynucleotide spanning at least one exon-exon junction identified in Table 1.

19) The detection reagent of claim 9, wherein said reagent is comprised of at least one polynucleotide spanning at least one exon-exon junction identified in Table 1.

20) The detection reagent of claim 13, wherein said reagent is comprised of at least one polynucleotide spanning at least one exon-exon junction identified in Table 1.

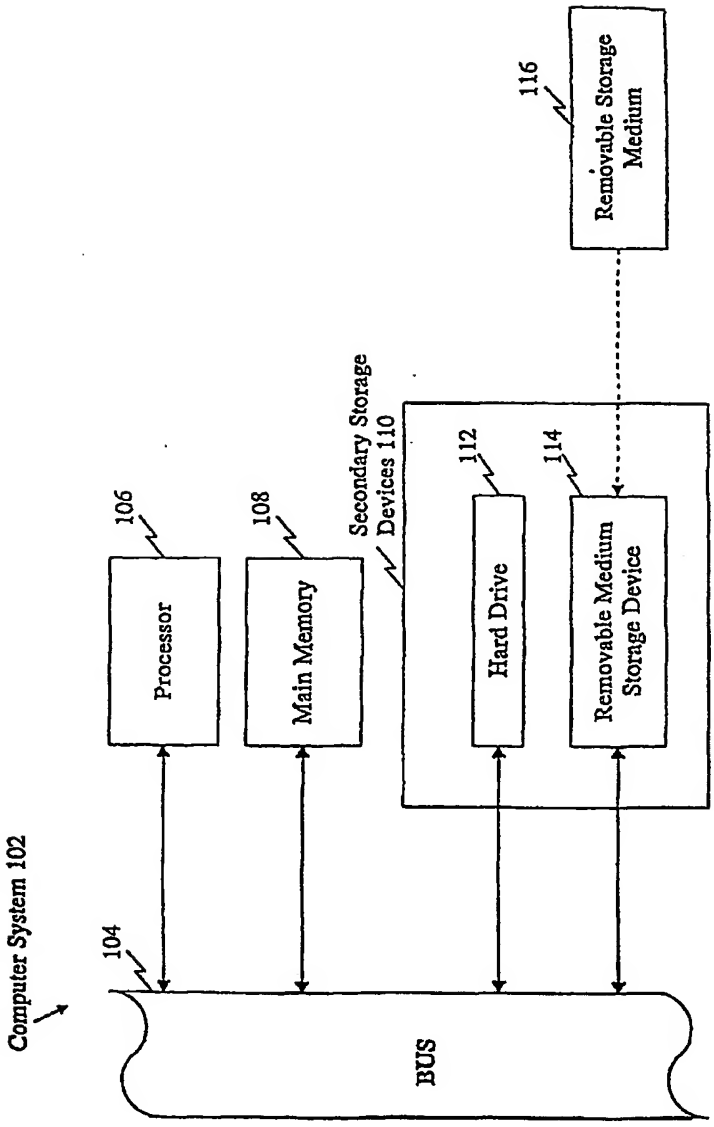

116
Removable Storage Medium
Secondary Storage Devices 110
112
114
106
108
Processor
Main Memory
Hard Drive
Removable Medium Storage Device
Computer System 102
104
BUS

FIGURE 1

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
6 September 2002 (06.09.2002)

PCT

(10) International Publication Number
WO 02/068579 A3

(51) International Patent Classification[7]: C12Q 1/68

(21) International Application Number: PCT/US02/00284

(22) International Filing Date: 10 January 2002 (10.01.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/756,696 10 January 2001 (10.01.2001) US

(71) Applicant: **PE CORPORATION (NY)** [US/US]; 761 Main Avenue, Norwalk, CT 06859 (US).

(72) Inventors: **VENTER, Craig, J.**; Celera Genomics, 45 West Gude Drive C2-4#21, Rockville, MD 20850 (US). **ADAMS, Mark**; Celera Genomics, 45 West Gude Drive C2-4#21, Rockville, MD 20850 (US). **LI, Peter, W., D.**; Celera Genomics, 45 West Gude Drive C2-4#21, Rockville, MD 20850 (US). **MYERS, Eugene, W.**; Celera Genomics, 45 West Gude Drive C2-4#21, Rockville, MD 20850 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

(88) Date of publication of the international search report:
31 July 2003

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/068579 A3

(54) Title: KITS, SUCH AS NUCLEIC ACID ARRAYS, COMPRISING A MAJORITY OF HUMAN EXONS OR TRANSCRIPTS, FOR DETECTING EXPRESSION AND OTHER USES THEREOF

(57) Abstract: The present invention is based on the sequencing and assembly of the human genome. The present invention provides the primary nucleotide sequence of the coding portion of the human genome in the form of a series of transcript sequences with accompanying exon information. This information can be used to generate nucleic acid detection reagents and kits such as nucleic acid arrays, and for other uses.

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 02/00284

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
BIOSIS, EPO-Internal, EMBASE, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	ZHU TONG ET AL: "Large-scale profiling of the Arabidopsis transcriptome." PLANT PHYSIOLOGY (ROCKVILLE), vol. 124, no. 4, December 2000 (2000-12), pages 1472-1476, XP002243507 ISSN: 0032-0889 page 1472, column 2 ---	1-20
Y	MIYAJIMA NOBUYUKI ET AL: "Computational and experimental analysis identifies many novel human genes." BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS, vol. 272, no. 3, 16 June 2000 (2000-06-16), pages 801-807, XP002243508 ISSN: 0006-291X the whole document --- -/--	1-20

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
5 June 2003	24/06/2003

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	CHAKRAVARTY, A

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 02/00284

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
X	BURGE C ET AL: "Prediction of complete gene structure in human genomic DNA" JOURNAL OF MOLECULAR BIOLOGY, LONDON, GB, vol. 268, no. 1, 25 April 1997 (1997-04-25), pages 78-94, XP002109301 ISSN: 0022-2836 the whole document	1-20
Y	WO 95 21265 A (SOUTHERN EDWIN MELLOR ;ISIS INNOVATION (GB); MIR KALIM ULLAH (GB);) 10 August 1995 (1995-08-10) claims	1-20
Y	SOLOVYEV VICTOR V ET AL: "Predicting internal exons by oligonucleotide composition and discriminant analysis of spliceable open reading frames." NUCLEIC ACIDS RESEARCH, vol. 22, no. 24, 1994, pages 5156-5163, XP002915964 ISSN: 0305-1048 abstract page 5156, column 2	17-20

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 02/00284

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☒ Claims Nos.: -
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
see FURTHER INFORMATION sheet PCT/ISA/210

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.2

Claims 1 is drafted in such a way as to attempt to define the subject-matter in terms of the result to be achieved (i.e. "capable of detecting the presence of 100,000 or more human exons"). In this instance the use of such a formulation renders the claims unclear in scope and is not justified by the disclosed means of achieving the desired result. The same objection applies to claims 2, 3, 5-7, 9-11, 13-15 and all dependent claims. (i.e. claim 1-20).
In the present case, the claims so lack support, and the application so lacks disclosure, that a meaningful search over the whole of the claimed scope is impossible.
Consequently, the search has been carried out for the concept of oligonucleotide arrays for detecting the presence of 100,000 or more human exons, in particular, using predicted transcript sequences and incorporating exon/exon junctions.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/US 02/00284

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9521265 A	10-08-1995	AT 202148 T	15-06-2001
		DE 69521298 D1	19-07-2001
		DE 69521298 T2	08-05-2002
		EP 0742837 A1	20-11-1996
		WO 9521265 A1	10-08-1995
		JP 9508791 T	09-09-1997
		US 2002031766 A1	14-03-2002
		US 6080585 A	27-06-2000

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
7 October 2004 (07.10.2004)

PCT

(10) International Publication Number
WO 2004/085648 A2

(51) International Patent Classification[7]: C12N 15/12, 15/62, 15/63, C07K 14/47, 16/18, A61K 38/17, 39/395, 48/00

(21) International Application Number: PCT/US2004/008323

(22) International Filing Date: 17 March 2004 (17.03.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/455,756	19 March 2003 (19.03.2003)	US
60/480,241	20 June 2003 (20.06.2003)	US
60/492,057	1 August 2003 (01.08.2003)	US

(71) Applicant *(for all designated States except US)*: **BIOGEN IDEC MA INC.** [US/US]; 14 Cambridge Center, Cambridge, MA 02142 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **MI, SHA** [US/US]; 4 Vernon Road, Belmont, MA 02478 (US). **MCCOY, JOHN** [GB/US]; 56 Howard Street, Reading, MA 01867 (US). **PEPINSKY, R., BLAKE** [US/US]; 30 Falmouth Road, Arlington, MA 02474 (US). **LEE, DANIEL, H., S.** [CA/US]; 361 Dutton Road, Sudbury, MA 01776 (US).

(74) Agents: **HALEY, JAMES, F., JR.** et al.; c/o Fish & Neave, 1251 Avenue of the Americas, New York, NY 10020 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

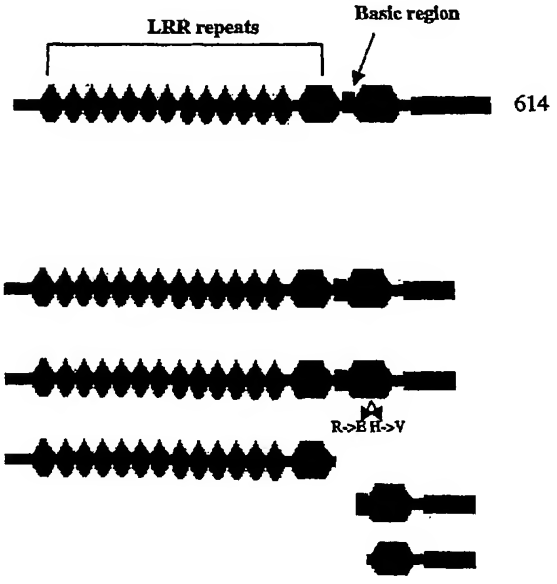
(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

[Continued on next page]

(54) Title: NOGO RECEPTOR BINDING PROTEIN



Deletion Construct (amino acids)	EC 50
(34-532)	6 nm
(34-532) mutated	120 nm
(34-416)	120 nm
(417-532)	60 nm
(425-532)	25 nm

(57) **Abstract:** The invention provides Sp35 polypeptides and fusion proteins thereof, Sp35 antibodies and antigen-binding fragments thereof and nucleic acids encoding the same. The invention also provides compositions comprising, and methods for making and using, such Sp35 antibodies, antigen-binding fragments thereof, Sp35 polypeptides and fusion proteins thereof.

WO 2004/085648 A2

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

NOGO RECEPTOR BINDING PROTEIN

Field of the Invention

[0001] This invention relates to neurology, neurobiology and molecular biology. More particularly, this invention relates to molecules and methods for treatment of neurological diseases, disorders and injuries such as spinal cord injury.

Background of the Invention

[0002] Axons and dendrites extend from neurons. The distal tip of an extending axon or neurite includes a specialized region, known as the growth cone. Growth cones sense the local environment and guide axonal growth toward a neuron's target cell. Growth cones respond to environmental cues, for example, surface adhesiveness, growth factors, neurotransmitters and electric fields. The growth cones generally advance at a rate of one to two millimeters per day. The growth cone explores the area ahead of it and on either side, by means of elongations classified as lamellipodia and filopodia. When an elongation contacts an unfavorable surface, it withdraws. When an elongation contacts a favorable growth surface, it continues to extend and guides the growth cone in that direction. When the growth cone reaches an appropriate target cell a synaptic connection is created.

[0003] Nerve cell function is influenced by contact between neurons and other cells in their immediate environment (Rutishauser, et al., 1988, Physiol. Rev. 68:819). These cells include specialized glial cells, oligodendrocytes in the central nervous system (CNS), and Schwann cells in the peripheral nervous system (PNS),

which sheathe the neuronal axon with myelin (Lemke, 1992, in An Introduction to Molecular Neurobiology, Z. Hall, Ed., p. 281, Sinauer).

[0004] CNS neurons have the inherent potential to regenerate after injury, but they are inhibited from doing so by inhibitory proteins present in myelin (Brittis et al., 2001, Neuron 30:11-14; Jones et al, 2002, J. Neurosci. 22:2792-2803; Grimpe et al, 2002, J. Neurosci.:22:3144-3160).

[0005] Several myelin inhibitory proteins found on oligodendrocytes have been characterized. Known examples of myelin inhibitory proteins include NogoA (Chen et al., Nature, 2000, 403, 434-439; Grandpre et al., Nature 2000, 403, 439-444), myelin associated glycoprotein (MAG) (McKerracher et al., 1994, Neuron 13:805-811; Mukhopadhyay et al., 1994, Neuron 13:757-767) and oligodendrocyte glycoprotein (OM-gp), Mikol et al., 1988, J. Cell. Biol.106:1273-1279). Each of these proteins has been separately shown to be a ligand for the neuronal NgR1 (Wang et al., Nature 2002, 417, 941-944; Grandpre et al., Nature 2000, 403, 439-444; Chen et al., Nature, 2000, 403, 434-439; Domeniconi et al., Neuron 2002, published online June 28, 2002).

[0006] Nogo receptor-1 (NgR1) is a GPI-anchored membrane protein that contains 8 leucine rich repeats (Fournier et al., 2001, Nature 409:341-346). Upon interaction with inhibitory proteins (e.g., NogoA, MAG and OM-gp), the NgR1 complex transduces signals that lead to growth cone collapse and inhibition of neurite outgrowth.

[0007] There is an unmet need for molecules and methods for inhibiting NgR1-mediated growth cone collapse and the resulting inhibition of neurite outgrowth.

Summary of the Invention

[0008] We have made various discoveries regarding a polypeptide designated "Sp35" (our designation). Alternate designations for Sp35 include "LINGO" and "LINGO-1." Our discoveries include the following. Sp35 binds to NgR1. Sp35 binds to itself in a homotypic interaction. An Sp35-Fc fusion protein induces or promotes fasciculation in granular neurons. An Sp35-Fc fusion protein promotes neuronal survival in both the rubro-spinal tract hemisection injury model and the optic nerve transection model. Sp35 retrovirus-infected cortical primary cells,

when delivered into spinal cord-injured rats, result in enhanced neuron survival, increased ß III tubulin staining of axons, and increased myelin content.

[0009] Based in part on these discoveries, the invention features an isolated nucleic acid containing a nucleotide sequence encoding a polypeptide wherein: (a) the polypeptide includes (i) an Sp35 LRR domain, (ii) an Sp35 basic region C-terminal to the LRR domain, and (iii) an Sp35 immunoglobulin (Ig) domain C-terminal to the basic region; and (b) the polypeptide lacks a transmembrane domain. The Sp35 LRR domain can contain a carboxy-terminal LRR (LRRCT), an amino-terminal LRR (LRRNT), or both. In some embodiments of the invention, the encoded Sp35 polypeptide lacks the cytoplasmic domain. In some embodiments, the encoded Sp35 polypeptide includes amino acid residues 34-532 of SEQ ID NO: 2 and lacks amino acid residues 533-614.

[0010] The invention also includes a nucleic acid encoding a polypeptide wherein the polypeptide includes an Sp35 Ig domain and lacks an Sp35 LRR domain, an Sp35 basic region, a transmembrane domain, and a cytoplasmic domain.

[0011] The invention also includes a nucleic acid encoding a polypeptide wherein the polypeptide includes an Sp35 LRR domain and lacks an Sp35 Ig domain, an Sp35 basic region, a transmembrane domain, and a cytoplasmic domain.

[0012] The invention also includes a nucleic acid encoding a polypeptide lacking a functional cytoplasmic domain but including all the other Sp35 domains. For example, the encoded polypeptide could include amino acids 1-576 of SEQ ID NO: 2 (prior to processing of the signal sequence).

[0013] In some embodiments of the invention, the encoded polypeptide is a fusion polypeptide containing a non-Sp35 moiety. The non-Sp35 moiety can be, for example, an Ig moiety, a serum albumin moiety, a targeting moiety, a reporter moiety, or a purification-facilitating moiety. A preferred non-Sp35 moiety is an Ig moiety, e.g., an Fc moiety.

[0014] The nucleotide sequence can be operatively linked to an expression control sequence, for example, in an expression vector. The invention also

includes a host cell transformed with a vector that expresses an Sp35 polypeptide of the invention.

[0015] The invention also includes an Sp35 polypeptide encoded by any of the above-described nucleic acids.

[0016] The invention also includes an Sp35 polypeptide conjugated to a polymer, e.g., a polyalkylene glycol, a sugar polymer, and a polypeptide. A preferred polymer is a polyalkylene glycol, e.g., polyethylene glycol (PEG). The polypeptide can be conjugated to 1, 2, 3 or 4 polymers. Preferably, the total molecular weight of the conjugated polymers is from 20,000 Da to 40,000 Da per Sp35 polypeptide.

[0017] The invention also includes a method of inhibiting signal transduction by NgR1. The method includes contacting the NgR1 with an effective amount of an Sp35 polypeptide. Preferred polypeptides for use in the method include the following:

(a) an Sp35 polypeptide, wherein: (a) the polypeptide includes (i) an Sp35 LRR domain, (ii) an Sp35 basic region C-terminal to the LRR domain, and (iii) an Sp35 immunoglobulin (Ig) domain C-terminal to the basic region; and (b) the polypeptide lacks a transmembrane domain; and

(b) an Sp35 polypeptide that includes an Sp35 Ig domain and lacks an Sp35 LRR domain, an Sp35 basic region, a transmembrane domain, and a cytoplasmic domain.

[0018] The invention also includes a method of decreasing inhibition of axonal growth of a central nervous system (CNS) neuron. The method includes contacting the neuron with an effective amount of a polypeptide such as an Sp35 polypeptide, an anti-Sp35 antibody, or an antigen-binding fragment of an anti-Sp35 antibody.

[0019] The invention also includes a method of inhibiting growth cone collapse of a CNS neuron. The method includes contacting the neuron with an effective amount of a polypeptide such as an Sp35 polypeptide, an anti-Sp35 antibody, or an antigen-binding fragment of an anti-Sp35 antibody.

[0020] The invention also includes a method of treating a CNS disease, disorder or injury in a mammal. The method includes administering to the mammal a

therapeutically effective amount of a polypeptide such as an Sp35 polypeptide, an anti-Sp35 antibody, or an antigen-binding fragment of an anti-Sp35 antibody. In some embodiments of the invention, the CNS disease, disorder or injury is a spinal cord injury. The Sp35 polypeptide can be administered locally. In some embodiments of the method, the Sp 35 polypeptide is administered initially within 48 hours of a spinal cord injury. For local administration, the therapeutically effective amount of the polypeptide preferably is from 10 μg/kg to 10 mg/kg. For systemic administration, the therapeutically effective amount of the polypeptide preferably is from 1 mg/kg to 20 mg/kg.

[0021] The invention also includes an *ex vivo* gene therapy method of treating a CNS disease, disorder or injury in a mammal. The method includes (a) providing a cultured host cell expressing a recombinant Sp35 polypeptide; and (b) introducing the host cell into the mammal at the site of the CNS disease, disorder or injury, e.g., spinal cord injury. The cultured host cell can be derived from the mammal to be treated. In this *ex vivo* gene therapy method, the recombinant Sp35 polypeptide can be a full-length Sp35 polypeptide.

[0022] The invention also includes a method of promoting myelination at the site of the CNS disease, disorder or injury. The method includes contacting the site of the CNS disease, disorder or injury with an effective amount of an Sp35 polypeptide, e.g., a polypeptide containing an Sp35 LRR domain and lacking an Sp35 Ig domain, an Sp35 basic region, a transmembrane domain, and a cytoplasmic domain.

[0023] The invention also includes an *in vivo* gene therapy method of treating a CNS disease, disorder or injury by *in vivo* gene therapy. The method includes the steps of administering to a mammal, at or near the site of the disease, disorder or injury, a viral vector containing a nucleotide sequence that encodes an Sp35 polypeptide so that the Sp35 polypeptide is expressed from the nucleotide sequence in the mammal in an amount sufficient to reduce inhibition of axonal extension by neurons at or near the site of the injury. The viral vector can be, e.g., an adenoviral vector, a lentiviral vector, a baculoviral vector, an Epstein Barr viral vector, a papovaviral vector, a vaccinia viral vector, and a herpes simplex viral vector. The disease, disorder or injury can be, e.g., spinal cord injury or optic

nerve injury. The viral vector can be administered by a route such as topical administration, intraocular administration, parenteral administration, intrathecal administration, subdural administration and subcutaneous administration.

[0024] The invention also includes a method of promoting survival of a neuron at risk of dying. The method includes contacting the neuron with an effective amount of an Sp35 polypeptide. The Sp35 polypeptide can be a soluble form of Sp35, e.g., an Sp35-Fc fusion protein. The neuron can be *in vitro* or *in vivo*, e.g., in a mammal with a neurodegenerative disease disorder or injury, e.g., multiple sclerosis, ALS, Huntington's disease, Alzheimer's disease, Parkinson's disease, diabetic neuropathy, stroke, traumatic brain injuries and spinal cord injury. In some embodiments of the invention, the Sp35 polypeptide is administered indirectly by: (a) providing a cultured host cell expressing a recombinant Sp35 polypeptide; and (b) introducing the host cell into the mammal at the site of the neuron. In some embodiments of the invention, the polypeptide is administered indirectly through *in vivo* gene therapy. In such an embodiment, the method includes administering, at or near the site of the neuron, a viral vector comprising a nucleotide sequence that encodes an Sp35 polypeptide so that the Sp35 polypeptide is expressed from the nucleotide sequence in the mammal in an amount sufficient to promote survival of the neuron.

[0025] As used herein, "full length human Sp35 polypeptide" means the polypeptide whose amino acid sequence is amino acids 34-614 of SEQ ID NO: 2.

[0026] As used herein, "heterologous moiety" means an amino acid sequence not present in a full-length Sp35 polypeptide.

[0027] As used herein, "nogo receptor-1" means the polypeptide whose sequence is publicly available under Genbank accession no. AAG53612.

[0028] As used herein, "Sp35 antagonist polypeptide" means an Sp35 polypeptide that blocks, inhibits, or interferes with the biological activity of naturally-occurring Sp35.

[0029] As used herein, "Sp35 basic region" means the following amino acid motif:

R	**R**	**A**	**R**	**I**	**R**	**D**	**R**	**K**	(SEQ ID NO: 4)
K	K	V	K	V	K	E	K	R	(SEQ ID NO: 5)
R	R	L	R	L	R	D	R	K	(SEQ ID NO: 6)
R	R	G	R	G	R	D	R	K	(SEQ ID NO: 7)
R	R	I	R	A	R	D	R	K	(SEQ ID NO: 8)

The top row of amino acids (in bold; SEQ ID NO: 4) is the preferred Sp35 basic region sequence, with variants showing optional substitutions shown below (SEQ ID NOS: 5, 6, 7 and 8).

[0030] As used herein, "Sp35 fusion protein" means a fusion protein that includes an Sp35 moiety fused to a heterologous moiety.

[0031] As used herein, "Sp35 Ig domain" means amino acids 433-493 of SEQ ID NO: 2, provided that the sequence can contain up to five individual amino acid insertions, deletions, or conservative amino acid substitutions. The following substitutions (numbering based on SEQ ID NO: 2) are expressly included: V to M at position 6; S to G at position 294; V to A at position 348; and R to H at position 419.

[0032] As used herein, "Sp35 LRR domain" means a domain that includes 10 to 14 of the leucine rich repeat sequences, including the LRRNT and LRRCT, listed in Table 1, provided that up to five amino acid insertions, deletions, or conservative amino acid substitutions can appear within the aggregate 10-14 leucine rich repeats.

[0033] As used herein, "Sp35 moiety" means a biologically active fragment of a full-length Sp35 polypeptide.

[0034] As used herein, "Sp35 polypeptide" means an Sp35 moiety or a fusion protein that includes an Sp35 moiety.

[0035] Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which the invention pertains. In case of conflict, the present specification, including definitions, will control. All publications, patents and other references mentioned herein are incorporated by reference.

[0036] Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the invention, the preferred methods and materials are described below. The materials, methods and examples are illustrative only, and are not intended to be limiting. Other features and advantages of the invention will be apparent from the detailed description and from the claims.

Brief Description of the Drawings

[0037] FIG. 1 is the nucleotide sequence of a full-length human Sp35 cDNA (SEQ ID NO:1)

[0038] FIG. 2 is the amino acid sequence of a full-length human Sp35 polypeptide (SEQ ID NO: 2).

[0039] FIG. 3 is a schematic illustration of the Sp35 domain structure and deletion mapping to identify Sp35 sequence(s) that bind to NgR1.

[0040] FIG. 4 is a histogram summarizing data on SP35 binding to COS7 cells transfected with an expression vector encoding rat p75 or a vector control. After 48 hours, AP-SP35 or AP was incubated with the cells. Bound AP was detected using chromogenic AP detection reagent.

[0041] FIG. 5 is a histogram summarizing data on binding of AP-Omgp, and AP-Nogo-66 to COS7 cells transfected with an expression vector encoding NgR1; NgR1 and p75; NgR1, p75, and SP35, or a vector control. After 48 hours, AP-Omgp, AP-Nogo-66 or AP was incubated with the cells. Bound AP was detected using chromogenic AP detection reagent.

[0042] FIG. 6 is a histogram summarizing data on the relief of inhibitory activity of myelin inhibitors on neurite outgrowth *in vitro*. Neurite length was measured on postnatal day 7 rat cerebellar granular neurons expressing DN-Sp35, full-length Sp35, or controls cultured on immobolized substrate Omgp, Myelin and Nogo-66. DN-SP35 transfected cells exhibited diminished response to inhibitory substrates. Neurite length was quantified from 1000 neurons per treatment group from two independent experiments ($p < 0.01$).

[0043] FIG. 7 is a histogram summarizing data on reversal of inhibitory activity of myelin inhibitors by SP35-Fc. Neurite length of postnatal day 7 rat cerebellar granular neurons (1000 neurons) cultured on immobilized substrate OMgp, Myelin or Nogo-66 in the presence or absence of SP35-Fc. SP35-Fc reduced the neurite

outgrowth inhibition caused by Omgp, Nogo-66 and MAG. Neurite length was quantified from 1000 neurons per treatment group from two independent experiments ($p < 0.01$).

[0044] FIG. 8 is a graph summarizing data from an experiment showing that intrathecal-administered Sp35-Fc improves functional recovery after dorsal hemisection in a rat. The locomoter BBB score measured as a function of time after dorsal hemisection in control (IgG) or Sp35-Fc-treated rats (8 animals per group). Treatment was initiated at the time of spinal cord injury.

[0045] FIG. 9 is a graph showing individual animal BBB scores at week four in the experiment summarized in FIG. 8.

Detailed Description of the Invention

[0046] Naturally occurring human Sp35 is a glycosylated CNS-specific protein containing 614 amino acids (FIG. 2; SEQ ID NO: 2). The human, full-length wild-type SP35 polypeptide contains an LRR domain consisting of 14 leucine-rich repeats (including N- and C-terminal caps), an Ig domain, a transmembrane region, and a cytoplasmic domain (FIG. 3). The cytoplasmic domain contains a canonical tyrosine phosphorylation site. In addition, the naturally occurring Sp35 protein contains a signal sequence, a short basic region between the LRRCT and Ig domain, and a transmembrane region between the Ig domain and the cytoplasmic domain (FIG. 3). The human Sp35 gene contains alternative translation start codons, so that six additional amino acids, i.e., MQVSKR (SEQ ID NO: 9) may or may not be present at the N-terminus of the Sp35 signal sequence. Table 1 lists the Sp35 domains and other regions, according to amino acid residue number, based on the sequence in FIG. 2 (SEQ ID NO: 2).

Table 1

Domain or Region	Beginning Residue	Ending Residue
Signal Sequence	1	33
LRRNT	34	64
LRR	66	89
LRR	90	113
LRR	114	137
LRR	138	161
LRR	162	185
LRR	186	209
LRR	210	233
LRR	234	257
LRR	258	281
LRR	282	305
LRR	306	329
LRR	330	353
LRRCT	363	416
Basic	417	424
Ig	433	493
Connecting sequence	494	551
Transmembrane	552	576
Cytoplasmic	577	614

[0047] Tissue distribution and developmental expression of Sp35 has been studied in humans and rats. Sp35 biology has been studied in an experimental animal (rat) model. Expression of rat SP35 is localized to CNS neurons and brain oligodendrocytes, as determined by northern blot and immuno-histochemical staining. Rat Sp35 mRNA expression level is regulated developmentally, peaking shortly after birth, i.e., ca. postnatal day one. In a rat spinal cord transection injury model, Sp35 is up-regulated at the injury site, as determined by RT-PCR.

[0048] The inventors have discovered that full-length, wild-type Sp35 binds to NgR1. Soluble derivatives of Sp35 function as Sp35 antagonist polypeptides by

binding to NgR1 and blocking, inhibiting, or interfering with its function, thereby relieving the NgR1-mediated inhibition of axonal extension that normally takes place in CNS neurons. This is beneficial in situations where axonal extension or neurite sprouting is needed in the brain or spinal cord. Spinal cord injury, including partial or complete crush or severance, exemplifies a situation in which axonal extension is needed, but is normally inhibited through operation of the Nogo pathway. Examples of diseases or disorders in which axonal extension and/or neurite sprouting in the brain would be beneficial include stroke, multiple sclerosis, and other neurodegenerative diseases or disorders.

[0049] In methods of the present invention, an Sp35 polypeptide or an Sp35 blocking antibody (or antigen-binding antibody fragment) can be administered directly as a preformed polypeptide, or indirectly through a nucleic acid vector, to antagonize NgR1 function and permit beneficial axonal outgrowth.

[0050] In some embodiments of the invention a soluble Sp35 antagonist polypeptide is administered in a treatment method that includes: (1) transforming or transfecting an implantable host cell with a nucleic acid, e.g., a vector, that expresses an Sp35 polypeptide; and (2) implanting the transformed host cell into a mammal, at the site of a disease, disorder or injury. For example, the transformed host cell can be implanted at the site of a spinal cord injury. In some embodiments of the invention, the implantable host cell is removed from a mammal, temporarily cultured, transformed or transfected with an isolated nucleic acid encoding a soluble Sp35 polypeptide, and implanted back into the same mammal from which it was removed. The cell can be, but is not required to be, removed from the same site at which it is implanted. Such embodiments, sometimes known as *ex vivo* gene therapy, can provide a continuous supply of the Sp35 polypeptide, localized at the site of site of action, for a limited period of time.

[0051] The invention provides oligopeptides useful as modulators of the Sp35 interaction with NgR1 and Sp35 homotypic interactions. The oligopeptides include the following amino acid motif:

L S P R K H		(SEQ ID NO: 10)
I T P K R R		(SEQ ID NO: 11)
A C P H H K		(SEQ ID NO: 12)
V S P R K H		(SEQ ID NO: 13)

[0052] The top row of amino acids (in bold; SEQ ID NO: 10) is the preferred sequence, with variants comprising optional substitutions shown below (SEQ ID NOS: 11, 12 and 13).

[0053] Various exemplary Sp35 polypeptides, anti-Sp35 antibodies and antibody fragments, and methods and materials for obtaining these molecules for practicing the present invention are described below.

Fusion Proteins and Conjugated Polypeptides

[0054] Some embodiments of the invention involve the use of an Sp35 polypeptide, e.g., an Sp35 antagonist polypeptide, wherein an Sp35 moiety is fused to a heterologous polypeptide moiety to form an Sp35 fusion protein. Sp35 fusion proteins can be used to accomplish various objectives, e.g., increased serum half-life, improved bioavailability, *in vivo* targeting to a specific organ or tissue type, improved recombinant expression efficiency, improved host cell secretion, ease of purification, and higher avidity. Depending on the objective(s) to be achieved, the heterologous moiety can be inert or biologically active. Also, it can be chosen to be stably fused to the Sp35 moiety or to be cleavable, *in vitro* or *in vivo*. Heterologous moieties to accomplish different objectives are known in the art.

[0055] As an alternative to expression of a Sp35 fusion protein, a chosen heterologous moiety can be preformed and chemically conjugated to the Sp35 moiety. In most cases, a chosen heterologous moiety will function similarly, whether fused or conjugated to the Sp35 moiety. Therefore, in the following discussion of heterologous amino acid sequences, unless otherwise noted, it is to be understood that the heterologous sequence can be joined to the Sp35 moiety in the form of a fusion protein or as a chemical conjugate.

[0056] Pharmacologically active polypeptides such as Sp35 often exhibit rapid *in vivo* clearance, necessitating large doses to achieve therapeutically effective concentrations in the body. In addition, polypeptides smaller than about 60 kDa potentially undergo glomerular filtration, which sometimes leads to nephrotoxicity. Fusion or conjugation of relatively small polypeptides such as Sp35 fragments can be employed to reduce or avoid the risk of such nephrotoxicity. Various heterologous amino acid sequences, i.e., polypeptide moieties or "carriers," for increasing the *in vivo* stability, i.e., serum half-life, of therapeutic polypeptides are known.

[0057] Due to its long half-life, wide *in vivo* distribution, and lack of enzymatic or immunological function, essentially full-length human serum albumin (HSA), or an HSA fragment, is a preferred heterologous moiety. Through application of methods and materials such as those taught in Yeh et al., 1992, Proc. Natl. Acad. Sci. USA, 89:1904-1908 and Syed et al., 1997, Blood 89:3243-3252, HSA can be used to form an Sp35 fusion protein or conjugate that displays pharmacological activity by virtue of the Sp35 moiety while displaying significantly increased, e.g., 10-fold to 100-fold higher, *in vivo* stability. Preferably, the C-terminus of the HSA is fused to the N-terminus of the Sp35 moiety. Since HSA is a naturally secreted protein, the HSA signal sequence can be exploited to obtain secretion of the Sp35 fusion protein into the cell culture medium, when the fusion protein is produced in a eukaryotic, e.g., mammalian, expression system.

[0058] Some embodiments of the invention employ an Sp35 polypeptide wherein an Sp35 moiety is fused to an Fc region, i.e., the C-terminal portion of an Ig heavy chain constant region. Potential advantages of an Sp35-Fc fusion include solubility, *in vivo* stability, and multivalency, e.g., dimerization. The Fc region used can be an IgA, IgD, or IgG Fc region (hinge-CH2-CH3). Alternatively, it can be an IgE or IgM Fc region (hinge-CH2-CH3-CH4). An IgG Fc region is preferred, e.g., an IgG1 Fc region or IgG4 Fc region. Materials and methods for constructing and expressing DNA encoding Fc fusions are known in the art and can be applied to obtain Sp35 fusions without undue experimentation. Some embodiments of the invention employ an Sp35 fusion protein such as those described in Capon et al. U.S. Patent Nos. 5,428,130 and 5,565,335.

[0059] The signal sequence is a polynucleotide that encodes an amino acid sequence that initiates transport of a protein across the membrane of the endoplasmic reticulum. Signal sequences useful for constructing an immunofusin include antibody light chain signal sequences, e.g., antibody 14.18 (Gillies et.al., 1989, J. Immunol. Meth., 125:191-202), antibody heavy chain signal sequences, e.g., the MOPC141 antibody heavy chain signal sequence (Sakano et al., 1980, Nature 286:5774). Alternatively, other signal sequences can be used. See, for example, Watson, 1984, Nucleic Acids Research 12:5145). The signal peptide is usually cleaved in the lumen of the endoplasmic reticulum by signal peptidases. This results in the secretion of a immunofusin protein containing the Fc region and the Sp35 moiety.

[0060] In some embodiments the DNA sequence encodes a proteolytic cleavage site between the secretion cassette and the Sp35 moiety. A cleavage site provides for the proteolytic cleavage of the encoded fusion protein, thus separating the Fc domain from the target protein. Useful proteolytic cleavage sites include amino acids sequences recognized by proteolytic enzymes such as trypsin, plasmin, thrombin, factor Xa, or enterokinase K.

[0061] The secretion cassette can be incorporated into a replicable expression vector. Useful vectors include linear nucleic acids, plasmids, phagemids, cosmids and the like. An exemplary expression vector is pdC, in which the transcription of the immunofusin DNA is placed under the control of the enhancer and promoter of the human cytomegalovirus. See, e.g., Lo et al., 1991, Biochim. Biophys. Acta 1088:712; and Lo et al., 1998, Protein Engineering 11:495-500. An appropriate host cell can be transformed or transfected with a DNA that encodes an Sp35 polypeptide, and is used for the expression and secretion of the Sp35 polypeptide. Preferred host cells include immortal hybridoma cells, myeloma cells, 293 cells, Chinese hamster ovary (CHO) cells, Hela cells, and COS cells.

[0062] Fully intact, wild-type Fc regions display effector functions that normally are unnecessary and undesired in an Fc fusion protein according to the present invention. Therefore, certain binding sites preferably are deleted from the Fc region during the construction of the secretion cassette. For example, since coexpression with the light chain is unnecessary, the binding site for the heavy

chain binding protein, Bip (Hendershot et al., 1987, Immunol. Today 8:111-114), is deleted from the CH2 domain of the Fc region of IgE, such that this site does not interfere with the efficient secretion of the immunofusin. Likewise, the cysteine residues present in the Fc regions which are responsible for binding to the light chain of the immunoglobulin should be deleted or substituted with another amino acid, such that these cysteine residues do not interfere with the proper folding of the Fc region when it is produced as an immunofusin. Transmembrane domain sequences, such as those present in IgM, should be deleted.

[0063] The IgG1 Fc region is preferred. Alternatively, the Fc region of the other subclasses of immunoglobulin gamma (gamma-2, gamma-3 and gamma-4) can be used in the secretion cassette. The IgG1 Fc region of immunoglobulin gamma-1 is preferably used in the secretion cassette includes the hinge region (at least part), the CH2 region, and the CH3 region. In some embodiments, the Fc region of immunoglobulin gamma-1 is a CH2-deleted-Fc, which includes part of the hinge region and the CH3 region, but not the CH2 region. A CH2-deleted-Fc has been described by Gillies et al., 1990, Hum. Antibod. Hybridomas, 1:47. In some embodiments, the Fc regions of IgA, IgD, IgE, or IgM, are used.

[0064] Sp35-Fc fusion proteins can be constructed in several different configurations. In one configuration the C-terminus of the Sp35 moiety is fused directly to the N-terminus of the Fc moiety. In a slightly different configuration, a short polypeptide, e.g., 2-10 amino acids, is incorporated into the fusion between the N-terminus of the Sp35 moiety and the C-terminus of the Fc moiety. Such a linker provides conformational flexibility, which may improve biological activity in some circumstances. If a sufficient portion of the hinge region is retained in the Fc moiety, the Sp35-Fc fusion will dimerize, thus forming a divalent molecule. A homogeneous population of monomeric Fc fusions will yield monospecific, bivalent dimers. A mixture of two monomeric Fc fusions each having a different specificity will yield bispecific, bivalent dimers.

[0065] Any of a number of cross-linkers that contain a corresponding amino reactive group and thiol reactive group can be used to link Sp35 to serum albumin. Examples of suitable linkers include amine reactive cross-linkers that insert a thiol reactive-maleimide, *e.g.*, SMCC, AMAS, BMPS, MBS, EMCS, SMPB, SMPH,

KMUS, and GMBS. Other suitable linkers insert a thiol reactive-haloacetate group, *e.g.*, SBAP, SIA, SIAB. Linkers that provide a protected or non-protected thiol for reaction with sulfhydryl groups to product a reducible linkage include SPDP, SMPT, SATA, and SATP. Such reagents are commercially available (e.g., Pierce Chemicals).

[0066] Conjugation does not have to involve the N-terminus of an Sp35 polypeptide or the thiol moiety on serum albumin. For example, Sp35-albumin fusions can be obtained using genetic engineering techniques, wherein the Sp35 moiety is fused to the serum albumin gene at its N-terminus, C-terminus, or both.

[0067] Sp35 polypeptides can be fused to heterologous peptides to facilitate purification or identification of the Sp35 moiety. For example, a histidine tag can be fused to an Sp35 polypeptide to facilitate purification using commercially available chromatography media.

[0068] In some embodiments of the invention, an Sp35 fusion construct is used to enhance the production of an Sp35 moiety in bacteria. In such constructs a bacterial protein normally expressed and/or secreted at a high level is employed as the N-terminal fusion partner of an Sp35 polypeptide. See, e.g., Smith et al., 1988 Gene 67:31; Hopp et al., 1988, Biotechnology 6:1204; La Vallie et al., 1993, Biotechnology 11:187.

[0069] In some embodiments of the invention, a fusion construct includes an Sp35 moiety and a second human NgR1-binding moiety, e.g., an oligodendrocyte-myelin glycoprotein (OMgp) moiety, a myelin associated glycoprotein (MAG) moiety, or Nogo66 moiety. Advantages of such constructs include increased NgR1 binding affinity.

[0070] The full-length OMgp amino acid sequence is known in the art (Genbank accession no. P23515). Specific examples of Sp35-OMgp fusions include the following:

[0071] Sp35 (aa 34-532) + IgG1 Fc + OMgp (amino acid residues 25-400); and

[0072] Sp35 (aa 34-532) + HSA + OMgp (amino acid residues 25-400).

[0073] The full-length MAG amino acid sequence is known in the art (Genbank accession no. A61084). Specific examples of Sp35-MAG fusions include the following:

[0074] Sp35 (aa 34-532) + IgG1 Fc + MAG (amino acid residues 12-500); and

[0075] Sp35 (aa 34-532) + HSA + MAG (amino acid residues 12-500).

[0076] The full-length Nogo amino acid sequence is known in the art (NogoA Genbank accession no. AY102279). Specific examples of Sp35-Nogo fusions include the following:

[0077] Sp35 (aa 34-532) + IgG1 Fc + Nogo66 (NogoA amino acid residues 1056-1122);

[0078] Sp35 (aa 34-532) + HSA + Nogo66 (NogoA amino acid residues 1056-1122);

[0079] Sp35 (aa 34-532) + IgG1 Fc + amino Nogo (NogoA amino acid residues 1-949); and

[0080] Sp35 (aa 34-532) + HSA + amino Nogo (NogoA amino acid residues 1-949).

[0081] By fusing an Sp35 moiety at the amino and carboxy termini of a suitable fusion partner, bivalent or tetravalent forms of an Sp35 polypeptide can be obtained. For example, an Sp35 moiety can be fused to the amino and carboxy termini of an Ig moiety to produce a bivalent monomeric polypeptide containing two Sp35 moieties. Upon dimerization of two of these monomers, by virtue of the Ig moiety, a tetravalent form of an Sp35 protein is obtained. Such multivalent forms can be used to achieve increased binding affinity for the target. Multivalent forms of Sp35 also can be obtained by placing Sp35 moieties in tandem to form concatamers, which can be employed alone or fused to a fusion partner such as Ig or HSA.

Conjugated Polymers (other than polypeptides)

[0082] Some embodiments of the invention involve an Sp35 polypeptide wherein one or more polymers are conjugated (covalently linked) to the Sp35 polypeptide. Examples of polymers suitable for such conjugation include polypeptides (discussed above), sugar polymers and polyalkylene glycol chains. Typically, but not necessarily, a polymer is conjugated to the Sp35 polypeptide for the purpose of improving one or more of the following: solubility, stability, or bioavailability.

[0083] A preferred class of polymer for conjugation to an Sp35 polypeptide is a polyalkylene glycol. Polyethylene glycol (PEG) is particularly preferred. PEG

moieties, e.g., 1, 2, 3, 4 or 5 PEG polymers, can be conjugated to each Sp35 polypeptide to increase serum half life, as compared to the Sp35 polypeptide alone. PEG moieties are non-antigenic and essentially biologically inert. PEG moieties used in the practice of the invention may be branched or unbranched.

[0084] The number of PEG moieties attached to the Sp35 polypeptide and the molecular weight of the individual PEG chains can vary. In general, the higher the molecular weight of the polymer, the fewer polymer chains attached to the polypeptide. Preferably, the total polymer mass attached to the Sp35 polypeptide is from 20 kDa to 40 kDa. Thus, if one polymer chain is attached, the preferred molecular weight of the chain is 20-40 kDa. If two chains are attached, the preferred molecular weight of each chain is 10-20 kDa. If three chains are attached, the preferred molecular weight is 7-14 kDa.

[0085] The polymer, e.g., PEG, can be linked to the Sp35 polypeptide through any suitable, exposed reactive group on the polypeptide. The exposed reactive group(s) can be, for example, an N-terminal amino group or the epsilon amino group of an internal lysine residue, or both. An activated polymer can react and covalently link at any free amino group on the Sp35 polypeptide. Free carboxylic groups, suitably activated carbonyl groups, hydroxyl, guanidyl, imidazole, oxidized carbohydrate moieties and mercapto groups of the Sp35 (if available) also can be used as reactive groups for polymer attachment.

[0086] Preferably, in a conjugation reaction, from about 1.0 to about 10 moles of activated polymer per mole of polypeptide, depending on polypeptide concentration, is employed. Usually, the ratio chosen represents a balance between maximizing the reaction while minimizing side reactions (often non-specific) that can impair the desired pharmacological activity of the Sp35 moiety. Preferably, at least 50% of the biological activity (as demonstrated, e.g., in any of the assays described herein or known in the art) of the Sp35 polypeptide is retained, and most preferably nearly 100% is retained.

[0087] The polymer can be conjugated to the Sp35 polypeptide using conventional chemistry. For example, a polyalkylene glycol moiety can be coupled to a lysine epsilon amino group of the Sp35 polypeptide. Linkage to the lysine side chain can be performed with an N-hydroxylsuccinimide (NHS) active

ester such as PEG succinimidyl succinate (SS-PEG) and succinimidyl propionate (SPA-PEG). Suitable polyalkylene glycol moieties include, *e.g.*, carboxymethyl-NHS, norleucine-NHS, SC-PEG, tresylate, aldehyde, epoxide, carbonylimidazole, and PNP carbonate. These reagents are commercially available. Additional amine reactive PEG linkers can be substituted for the succinimidyl moiety. These include, *e.g.*, isothiocyanates, nitrophenylcarbonates, epoxides, and benzotriazole carbonates. Conditions preferably are chosen to maximize the selectivity and extent or reaction. Such optimization of reaction conditions is within ordinary skill in the art.

[0088] PEGylation can be carried out by any of the PEGylation reactions known in the art. See, e.g., Focus on Growth Factors, 3: 4-10, 1992; published European patent applications EP 0 154 316 and EP 0 401 384. PEGylation may be carried out using an acylation reaction or an alkylation reaction with a reactive polyethylene glycol molecule (or an analogous reactive water-soluble polymer).

[0089] PEGylation by acylation generally involves reacting an active ester derivative of polyethylene glycol. Any reactive PEG molecule can be employed in the PEGylation. A preferred activated PEG ester is PEG esterified to N-hydroxysuccinimide (NHS). As used herein, "acylation" includes without limitation the following types of linkages between the therapeutic protein and a water soluble polymer such as PEG: amide, carbamate, urethane, and the like. See, Bioconjugate Chem. 5: 133-140, 1994. Reaction parameters should be chosen to avoid temperature, solvent, and pH conditions that would damage or inactivate the Sp35 polypeptide.

[0090] Preferably, the connecting linkage is an amide. Preferably, at least 95% of the resulting product is mono, di- or tri-PEGylated. However, some species with higher degrees of PEGylation may be formed in amounts depending on the specific reaction conditions used. Optionally, purified PEGylated species are separated from the mixture, particularly unreacted species, by conventional purification methods, including, e.g., dialysis, salting-out, ultrafiltration, ion-exchange chromatography, gel filtration chromatography, and electrophoresis.

[0091] PEGylation by alkylation generally involves reacting a terminal aldehyde derivative of PEG with Sp35 in the presence of a reducing agent. In addition, one

can manipulate the reaction conditions to favor PEGylation substantially only at the N-terminal amino group of Sp35 (i.e., a mono-PEGylated protein). In either case of mono-PEGylation or poly-PEGylation, the PEG groups are preferably attached to the protein via a -CH2-NH- group. With particular reference to the -CH2- group, this type of linkage is known as an "alkyl" linkage.

[0092] Derivatization via reductive alkylation to produce a mono-PEGylated product exploits differential reactivity of different types of primary amino groups (lysine versus the N-terminal) available for derivatization. The reaction is performed at a pH that allows one to take advantage of the pKa differences between the epsilon-amino groups of the lysine residues and that of the N-terminal amino group of the protein. By such selective derivatization, attachment of a water soluble polymer that contains a reactive group such as an aldehyde, to a protein is controlled: the conjugation with the polymer takes place predominantly at the N-terminus of the protein and no significant modification of other reactive groups, such as the lysine side chain amino groups, occurs.

[0093] The polymer molecules used in both the acylation and alkylation approaches are selected from among water soluble polymers. The polymer selected should be modified to have a single reactive group, such as an active ester for acylation or an aldehyde for alkylation, preferably, so that the degree of polymerization may be controlled as provided for in the present methods. An exemplary reactive PEG aldehyde is polyethylene glycol propionaldehyde, which is water stable, or mono C1-C10 alkoxy or aryloxy derivatives thereof (*see*, U.S. Patent 5,252,714). The polymer may be branched or unbranched. For the acylation reactions, the polymer(s) selected should have a single reactive ester group. For reductive alkylation, the polymer(s) selected should have a single reactive aldehyde group. Generally, the water soluble polymer will not be selected from naturally-occurring glycosyl residues since these are usually made more conveniently by mammalian recombinant expression systems.

[0094] Methods for preparing a PEGylated Sp35 generally includes the steps of (a) reacting a Sp35 protein or polypeptide with polyethylene glycol (such as a reactive ester or aldehyde derivative of PEG) under conditions whereby the molecule becomes attached to one or more PEG groups, and (b) obtaining the

reaction product(s). In general, the optimal reaction conditions for the acylation reactions will be determined case by case based on known parameters and the desired result. For example, the larger the ratio of PEG:protein, the greater the percentage of poly-PEGylated product.

[0095] Reductive alkylation to produce a substantially homogeneous population of mono- polymer/ Sp35 generally includes the steps of: (a) reacting a Sp35 protein or polypeptide with a reactive PEG molecule under reductive alkylation conditions, at a pH suitable to pen-nit selective modification of the N-terminal amino group of Sp35; and (b) obtaining the reaction product(s).

[0096] For a substantially homogeneous population of mono-polymer/ Sp35, the reductive alkylation reaction conditions are those that permit the selective attachment of the water soluble polymer moiety to the N-terminus of Sp35. Such reaction conditions generally provide for pKa differences between the lysine side chain amino groups and the N-terminal amino group. For purposes of the present invention, the preferred pH is in the range of 3-9, preferably 3-6.

[0097] Sp35 polypeptides can include a tag, *e.g.*, a moiety that can be subsequently released by proteolysis. Thus, the lysine moiety can be selectively modified by first reacting a His-tag modified with a low molecular weight linker such as Traut's reagent (Pierce) which will react with both the lysine and N-terminus, and then releasing the his tag. The polypeptide will then contain a free SH group that can be selectively modified with a PEG containing a thiol reactive head group such as a maleimide group, a vinylsulfone group, a haloacetate group, or a free or protected SH.

[0098] Traut's reagent can be replaced with any linker that will set up a specific site for PEG attachment. For example, Traut's reagent can be replaced with SPDP, SMPT, SATA, or SATP (Pierce). Similarly one could react the protein with an amine reactive linker that inserts a maleimide (for example SMCC, AMAS, BMPS, MBS, EMCS, SMPB, SMPH, KMUS, or GMBS), a haloacetate group (SBAP, SIA, SIAB), or a vinylsulfone group and react the resulting product with a PEG that contains a free SH.

[0099] In some embodiments, the polyalkylene glycol moiety is coupled to a cysteine group of the Sp35 polypeptide. Coupling can be effected using, *e.g.*, a maleimide group, a vinylsulfone group, a haloacetate group, or a thiol group.

[0100] Optionally, the Sp35 polypeptide is conjugated to the polyethylene glycol moiety through a labile bond. The labile bond can be cleaved in, *e.g.*, biochemical hydrolysis, proteolysis, or sulfhydryl cleavage. For example, the bond can be cleaved under *in vivo* (physiological) conditions.

[0101] The reactions may take place by any suitable method used for reacting biologically active materials with inert polymers, preferably at about pH 5-8, *e.g.*, pH 5, 6, 7, or 8, if the reactive groups are on the alpha amino group at the N-terminus. Generally the process involves preparing an activated polymer and thereafter reacting the protein with the activated polymer to produce the soluble protein suitable for formulation.

Vectors

[0102] The invention provides vectors comprising the nucleic acids encoding Sp35 polypeptides. The choice of vector and expression control sequences to which the nucleic acids of this invention is operably linked depends on the functional properties desired, *e.g.*, protein expression, and the host cell to be transformed.

[0103] Expression control elements useful for regulating the expression of an operably linked coding sequence are known in the art. Examples include, but are not limited to, inducible promoters, constitutive promoters, secretion signals, and other regulatory elements. When an inducible promoter is used, it can be controlled, e.g., by a change in nutrient status, or a change in temperature, in the host cell medium.

[0104] The vector can include a prokaryotic replicon, *i.e.*, a DNA sequence having the ability to direct autonomous replication and maintenance of the recombinant DNA molecule extra-chromosomally in a bacterial host cell. Such replicons are well known in the art. In addition, vectors that include a prokaryotic replicon may also include a gene whose expression confers a detectable marker such as a drug resistance. Examples of bacterial drug resistance genes are those that confer resistance to ampicillin or tetracycline.

[0105] Vectors that include a prokaryotic replicon can also include a prokaryotic or bacteriophage promoter for directing expression of the coding gene sequences in a bacterial host cell. Promoter sequences compatible with bacterial hosts are typically provided in plasmid vectors containing convenient restriction sites for insertion of a DNA segment to be expressed. Examples of such plasmid vectors are pUC8, pUC9, pBR322 and pBR329 (BioRad), pPL and pKK223 (Pharmacia). Any suitable prokaryotic host can be used to express a recombinant DNA molecule encoding a protein of the invention.

[0106] Eukaryotic cell expression vectors are known in the art and are commercially available. Typically, such vectors contain convenient restriction sites for insertion of the desired DNA segment. Exemplary vectors include pSVL and pKSV-10 (Pharmacia), pBPV-1, pML2d (International Biotechnologies), pTDT1 (ATCC 31255), retroviral expression vector pMIG, adenovirus shuttle vector pDC315, and AAV vectors.

[0107] Eukaryotic cell expression vectors may include a selectable marker, e.g., a drug resistance gene. The neomycin phosphotransferase (neo) gene (Southern et al., 1982, J. Mol. Anal. Genet. 1:327-341) is an example of such a gene.

[0108] To express the antibodies or antibody fragments, DNAs encoding partial or full-length light and heavy chains are inserted into expression vectors such as plasmids, retroviruses, cosmids, YACs, EBV-derived episomes, and the like. The expression vector and expression control sequences are chosen to be compatible with the expression host cell used. The antibody light chain gene and the antibody heavy chain gene can be inserted into separate vector. In some embodiments, both genes are inserted into the same expression vector.

[0109] A convenient vector is one that encodes a functionally complete human C_H or C_L immunoglobulin sequence. Preferably, restriction sites engineered so that any V_H or V_L sequence can be easily inserted and expressed. In such vectors, splicing usually occurs between the splice donor site in the inserted J region and the splice acceptor site preceding the human C region, and also at the splice regions that occur within the human C_H exons. Polyadenylation and transcription termination occur at native chromosomal sites downstream of the coding regions.

The recombinant expression vector can also encode a signal peptide that facilitates secretion of the antibody chain from a host cell.

[0110] Preferred regulatory sequences for mammalian host cell expression include viral elements that direct high levels of protein expression in mammalian cells, such as promoters and enhancers derived from retroviral LTRs, cytomegalovirus (CMV) (such as the CMV promoter/enhancer), Simian Virus 40 (SV40) (such as the SV40 promoter/enhancer), adenovirus, (*e.g.*, the adenovirus major late promoter (AdMLP)), polyoma and strong mammalian promoters such as native immunoglobulin and actin promoters. For further description of viral regulatory elements, and sequences thereof, see *e.g.*, Stinski U.S. Pat. No. 5,168,062; Bell U.S. Pat. No. 4,510,245; and Schaffner U.S. Pat. No. 4,968,615.

[0111] The recombinant expression vectors may carry sequences that regulate replication of the vector in host cells (*e.g.*, origins of replication) and selectable marker genes. The selectable marker gene facilitates selection of host cells into which the vector has been introduced (see *e.g.*, Axel U.S. Pat. Nos. 4,399,216; 4,634,665 and 5,179,017). For example, typically the selectable marker gene confers resistance to drugs, such as G418, hygromycin or methotrexate, on a host cell into which the vector has been introduced. Preferred selectable marker genes include the dihydrofolate reductase (DHFR) gene (for use in dhfr- host cells with methotrexate selection/amplification) and the neo gene (for G418 selection).

[0112] Nucleic acid molecules encoding Sp35 polypeptides and anti-Sp35 antibodies, and vectors comprising these nucleic acid molecules, can be used for transformation of a suitable host cell. Transformation can be by any suitable method. Methods for introduction of exogenous DNA into mammalian cells are well known in the art and include dextran-mediated transfection, calcium phosphate precipitation, polybrene-mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the DNA into nuclei. In addition, nucleic acid molecules may be introduced into mammalian cells by viral vectors.

[0113] Transformation of host cells can be accomplished by conventional methods suited to the vector and host cell employed. For transformation of prokaryotic host cells, electroporation and salt treatment methods can be employed

(Cohen et al., 1972, Proc. Natl. Acad. Sci. USA 69:2110-2114). For transformation of vertebrate cells, electroporation, cationic lipid or salt treatment methods can be employed. See, e.g., Graham et al., 1973, Virology 52:456-467; Wigler et al., 1979, Proc. Natl. Acad. Sci. USA 76:1373-1376f.

[0114] Mammalian cell lines available as hosts for expression are known in the art and include many immortalized cell lines available from the American Type Culture Collection (ATCC). These include, *inter alia,* Chinese hamster ovary (CHO) cells, NSO, SP2 cells, HeLa cells, baby hamster kidney (BHK) cells, monkey kidney cells (COS), human hepatocellular carcinoma cells (*e.g.*, Hep G2), A549 cells, and a number of other cell lines.

[0115] Expression of polypeptides from production cell lines can be enhanced using known techniques. For example, the glutamine synthetase (GS) system is commonly used for enhancing expression under certain conditions. See, e.g., European Patent Nos. 0216846, 0256055, and 0323997 and European Patent Application No. 89303964.4.

Host Cells

[0116] Host cells can be prokaryotic or eukaryotic. Preferred eukaryotic host cells include, but are not limited to, yeast and mammalian cells, e.g., Chinese hamster ovary (CHO) cells (ATCC Accession No. CCL61), NIH Swiss mouse embryo cells NIH-3T3 (ATCC Accession No. CRL1658), and baby hamster kidney cells (BHK). Other useful eukaryotic host cells include insect cells and plant cells. Exemplary prokaryotic host cells are *E. coli* and *Streptomyces*.

Formulations

[0117] Compositions containing Sp35 polypeptides, anti-Sp35 antibodies, or antigen binding fragments of anti-Sp35 antibodies may contain suitable pharmaceutically acceptable carriers. For example, they may contain excipients and/or auxiliaries that facilitate processing of the active compounds into preparations designed for delivery to the site of action. Suitable formulations for parenteral administration include aqueous solutions of the active compounds in water-soluble form, for example, water-soluble salts. In addition, suspensions of the active compounds as appropriate oily injection suspensions may be

administered. Suitable lipophilic solvents or vehicles include fatty oils, for example, sesame oil, or synthetic fatty acid esters, for example, ethyl oleate or triglycerides. Aqueous injection suspensions may contain substances that increase the viscosity of the suspension include, for example, sodium carboxymethyl cellulose, sorbitol and dextran. Optionally, the suspension may also contain stabilizers. Liposomes also can be used to encapsulate the molecules of the invention for delivery into cells or interstitial spaces. Exemplary pharmaceutically acceptable carriers are physiologically compatible solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, water, saline, phosphate buffered saline, dextrose, glycerol, ethanol and the like. In some embodiments, the composition comprises isotonic agents, for example, sugars, polyalcohols such as mannitol, sorbitol, or sodium chloride. In some embodiments, the compositions comprise pharmaceutically acceptable substances such as wetting or minor amounts of auxiliary substances such as wetting or emulsifying agents, preservatives or buffers, which enhance the shelf life or effectiveness of the active ingredients.

[0118] Compositions of the invention may be in a variety of forms, including, for example, liquid (*e.g.*, injectable and infusible solutions), dispersions, suspensions, semi-solid and solid dosage forms. The preferred form depends on the mode of administration and therapeutic application.

[0119] The composition can be formulated as a solution, micro emulsion, dispersion, liposome, or other ordered structure suitable to high drug concentration. Sterile injectable solutions can be prepared by incorporating the active ingredient in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active ingredient into a sterile vehicle that contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying that yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution. The proper fluidity of a solution can be maintained, for example, by the use of a coating

such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prolonged absorption of injectable compositions can be brought about by including in the composition an agent that delays absorption, for example, monostearate salts and gelatin.

[0120] The active ingredient can be formulated with a controlled-release formulation or device. Examples of such formulations and devices include implants, transdermal patches, and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, for example, ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for the preparation of such formulations and devices are known in the art. See *e.g.*, Sustained and Controlled Release Drug Delivery Systems, J. R. Robinson, ed., Marcel Dekker, Inc., New York, 1978.

[0121] Injectable depot formulations can be made by forming microencapsulated matrices of the drug in biodegradable polymers such as polylactide-polyglycolide. Depending on the ratio of drug to polymer, and the nature of the polymer employed, the rate of drug release can be controlled. Other exemplary biodegradable polymers are polyorthoesters and polyanhydrides. Depot injectable formulations also can be prepared by entrapping the drug in liposomes or microemulsions.

[0122] Supplementary active compounds can be incorporated into the compositions. In some embodiments, an Sp35 polypeptide, anti-Sp35 antibody or fragment thereof is coadministered with an anti-NgR1 antibody, or an antigen-binding fragments thereof, or soluble NgR1 polypeptides or NgR1 fusion protein.

[0123] Dosage regimens may be adjusted to provide the optimum desired response. For example, a single bolus may be administered, several divided doses may be administered over time, or the dose may be proportionally reduced or increased as indicated by the exigencies of the therapeutic situation. It is advantageous to formulate parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. See, e.g., Remington's Pharmaceutical Sciences (Mack Pub. Co., Easton, PA 1980).

[0124] In addition to the active compound, the liquid dosage form may contain inert ingredients such as water, ethyl alcohol, ethyl carbonate, ethyl acetate, benzyl

alcohol, benzyl benzoate, propylene glycol, 1,3-butylene glycol, dimethylformamide, oils, glycerol, tetrahydrofurfuryl alcohol, polyethylene glycols, and fatty acid esters of sorbitan.

Gene Therapy

[0125] An Sp35 polypeptide can be produced *in vivo* in a mammal, e.g., a human patient, using a gene therapy approach to treatment of a CNS disease, disorder or injury in which reducing inhibition of axonal extension would be therapeutically beneficial. This involves administration of a suitable Sp35 polypeptide-encoding nucleic acid operably linked to suitable expression control sequences. Preferably, these sequences are incorporated into a viral vector. Suitable viral vectors for such gene therapy include adenoviral vectors, lentiviral vectors, baculoviral vectors, Epstein Barr viral vectors, papovaviral vectors, vaccinia viral vectors, herpes simplex viral vectors, and adeno associated virus (AAV) vectors. The viral vector can be a replication-defective viral vector. A preferred adenoviral vector has a deletion in its E1 gene or E3 gene. When an adenoviral vector is used, preferably the mammal is not exposed to a nucleic acid encoding a selectable marker gene.

Examples

[0126] The invention is further illustrated by the following experimental examples. The examples are provided for illustrative purposes only, and are not to be construed as limiting the scope or content of the invention in any way.

Example 1: Sp35 Expression Pattern

[0127] Expression of Sp35 in human tissues was evaluated by Northern blot analysis. Multiple tissue blots containing 12 human major tissues or 14 human CNS tissues were hybridized overnight at 68°C with P^{32} labeled Sp35 probe (nucleotides 150-450 of the Sp35 cDNA sequence). The blots were washed 3 times with 2x SSC, 0.5% SDS, then 3 times with 0.5x SSC, 0.1% SDS. The blot was then exposed to X-ray film and the mRNA levels visualized by autoradiography.

[0128] Sp35 was highly expressed in the human brain but not in heart, skeletal muscle, colon, thymus, spleen, kidney, liver, small intestine, placenta, lung and peripheral blood leukocytes. Sp35 was expressed in all brain tissues tested,

including tissues isolated from frontal cortex, posterior cortex, entorhinal cortex, hippocampus, olfactory bulb, striatum, thalamus, cerebellum, midbrains, pons, medulla and spinal cord. A gradient of gene expression along the rostral/cordal axis was observed for Sp35, with highest levels in the cortical cortex and lowest levels in the spinal cord.

[0129] Immunohistochemical (IHC) staining was used to determine if Sp35 is expressed in specific brain cells. 4% paraformaldehyde fixed rat brains, spinal cord sections, or primary granular neuron cultures were incubated with primary antibodies to Sp35, as indicated, followed by secondary antibodies conjugated to Alexa 480 or 590 (Molecular Probes Inc.). The sections were then mounted in Vectashield and visualized by fluorescence microscopy. The anti-Sp35 specific antibodies used for IHC were generated from a Fab phage display library using MorPhosys technology.

[0130] Sp35 is expressed specifically in neurons and oligodendrocytes, but not in astrocytes. This was determined in experiments in which Rat brain tissue sections were stained with various agents, including anti-astrocyte marker GFAP, an antibody to an oligodendrocyte marker (O4), and an antibody to the neuronal marker ßIII tubulin, all counterstained with an anti-Sp35 antibody. Oligodendrocytes and neurons were intensely stained with the anti-Sp35 antibody. No staining of astrocytes was observed.

[0131] As an independent confirmation of the expression pattern of Sp35, we performed semi quantitative RT-PCR using mRNA extracted (Ambion kit) from rat primary cell cultures of purified astrocytes, oligodendrocytes, and cerebellum granular neurons. Forward primer AAGGCCCAGCAGGTGTTTGTGGA (SEQ ID NO: 14) and reverse primer TACTCGATCTCGATGTTGTGCTTT (SEQ ID NO: 15) were used. Following 26 cycles, a strong band was observed in the mRNA from neurons, a distinct but weaker signal was detected in oligodendrocyte mRNA, and no band was observed in astrocytes.

Example 2: Sp35-Fc Fusion Protein

[0132] To study the biological function of Sp35, a construct was made fusing the extra-cellular portion of human Sp35 (residuse 1-531) to the hinge and Fc region of human IgG1. A partial coding sequence for human Sp35 was obtained by PCR

from clone 227.2 (Incyte) using the forward primer 5'CAGCAGGTCGACGCGGC CGCATGCTGGCGGGGGGCGT3' (SEQ ID NO: 16) and reverse primer 5'CAGCAGGTCGACCTCGCCCGGCTGGTTGG3' (SEQ ID NO: 17).

[0133] The blunt end PCR product was subcloned into the SrfI site of the PCR SCRIPT AMP vector (Stratagene) to create PCR SCRIPT AMP-sp35. A SalI fragment was isolated from PCR SCRIPT AMP-sp35 and subcloned into the PCRCAMP Ig vector (derivative of Stratagene vector PCR SCRIPT AMP wherein the Fc gamma sequence is subcloned as a SalI(5') to NotI(3') fragment), fusing the Sp35 signal sequence and ectodomain sequence (codons 1-531) in-frame with sequences encoding the hinge and Fc region of human Ig1. Correct isolates were identified, and a NotI fragment encompassing the Sp35 Fc fragment was subcloned into the single Not I cloning site of the 293E expression vector, CH274, a derivative of commercial expression vector REP4 (Invitrogen). The Sp35-Fc fusion encoded by the new vector, CH274/sp35-Fc , was confirmed by DNA sequencing as plasmid GT123.

[0134] Stable cell lines expressing Sp35-Fc fusion protein were generated by electroporation of CHO host cells DG44 with plasmid GT123. Transfected CHO cells were cultured in alpha minus MEM in the presence of 10% dialyzed serum and 4mM glutamine to select for nucleoside-independent growth. Fourteen days post-transfection, cells were fed fresh media. To screen for cells expressing Sp35-Fc, CHO cells were labeled with Phycoerythrin (PE)-labeled goat anti-human IgG (Jackson Labs) and subjected to high speed flow cytometry sorting in a FACS Mo-Flo (Cytomation). The cells that expressed the highest levels of Sp35-Ig were selected. These cells were expanded in culture for 7 days, then re-labeled and re-sorted. Cells expressing the highest levels of Sp35-Ig were isolated as individual clones in 96-well plates. These clones were grown for two weeks and then fed fresh media one day prior to FACS analysis to check for expression levels. Clones that expressed the highest levels of Sp35-Fc were expanded, and frozen cell banks were established. The cell lines were adapted to grow in suspension culture in the serum free media BCM16. The titer of Sp35-Fc produced by these clones was determined by growing cell lines at 37° C for 4-5 passages, then growing the cells to 50% maximal cell density and culturing them for 10-15days at 28°C until the

viable cell density dropped to 75%. At this time, the culture media were harvested, cleared of cells and debris by centrifugation, and the culture supernatants titered for Sp35-Fc levels by Western blot analysis using an anti-human Ig antibody (Jackson Lab) as the probe.

[0135] Sp35-Fc fusion protein was purified from the clarified culture medium as follows: 9 ml of 1M HEPES pH 7.5 was added to 900 ml of conditioned medium. The medium was batch loaded for 3 hr at 4°C onto 3 ml of Protein A Sepharose (Pharmacia). The resin was collected in a 1.5 cm (I.D.) column, and washed four times with 3 ml PBS, two times with 4 ml of PBS containing 800 mM NaCl, and then again with 3 mL of PBS. The Sp35-Fc was eluted from the column with 25 mM NaH_2PO_4 pH 2.8, 100 mM NaCl in 1.5 mL fractions and neutralized by adding 75 µL of 0.5 M NaH_2PO_4 pH 8.6. Peak protein-containing fractions were identified by absorbence at 280 nm, pooled, and subjected to further purification on a 1 mL Protein A column. Prior to loading, NaCl was added to 600 mM and HEPES pH 7.5 to 50 mM. The column was washed twice with 600 µL of 10 mM HEPES pH 7.5, 1 M NaCl , and then with 1 mL PBS. Sp35-Fc was eluted from the column with 25 mM NaH_2PO_4 pH 2.8, 100 mM NaCl, collecting 0.5 mL fractions, and neutralized by adding 25 µL of 0.5 M NaH_2PO_4 pH 8.6. Peak protein-containing fractions were identified by absorbance at 280 nm and pooled. By reducing SDS-PAGE, the Sp35-Ig migrated as a single band (>95% pure) with an apparent mass of 90 kDa. Under non-reducing conditions, the protein ran as a dimer with an approximate mass of 180 kDa. The purified Sp35-Fc was aliquoted and stored at –70°C. The NotI fragment of GT123, which contains Sp35 amino acids 1-531 and human IgG1 Fc, was subcloned into the PV90 vector NotI site to create DB002.

Example 3: His-AP-Sp35 Fusion Protein

[0136] To study and isolate the receptor for Sp35, the protein was expressed in COS7 and CHO cells as a His-tagged-alkaline phosphatase (His-AP) fusion protein. The plasmid was constructed as follows: The extracellular domain of Sp35 (a.a. 34-532) was PCR amplified using primers (forward) 5'-AATTAAGAATTCACGGGCTGCCCGCCCCGCTGCGAGT-3' (SEQ ID NO:

18), containing an Eco RI cleavage site (underlined), and (reverse) 5'-TATATTTCTAGATCACTCGCCCGGCTGGTTGGAGATGAAAGCGA-3' (SEQ ID NO: 19), containing an Xba I cleavage site (underlined). The PCR product was cleaved with Xba I, the resulting sticky-end filled in with T4 DNA polymerase, then digested with Eco RI and gel purified. The digested product was ligated into a Hind III-filled in/ EcoR I His-AP fragment from the His-AP-pcDNA 1.1 vector (Invitrogen). The His-AP –Sp35 fragment was digested with Hind III and Eco RI, filled in, then ligated into the Not I-filled site in vector pV90. The DNA sequence of the insert was confirmed by DNA sequencing.

[0137] COS7 cells were split the day before transfection. His-AP-Sp35 vector DNA (8 μg) was used to transfect $5x10^6$ cells using lipofectamine (Invitrogen). The conditioned medium was harvested 48 hr post transfection.

[0138] We developed a CHO cell line expressing the His-AP-Sp35 fusion protein using the pV90 plasmid. CHO host cells DG44 ($2x10^6$ cells) were transfected with 100 µg of plasmid by electroporation. Cells were cultured in alpha minus MEM in the presence of 10% dialyzed serum and 4mM glutamine to select for nucleoside-independent growth. Fourteen days post transfection cells were fed fresh media in anticipation of screening by FACS Mo-Flo (Cytomation) sorting. Transfected CHO cells were labeled with the mouse monoclonal antibody 8B6 directed against human placental alkaline phosphatase (Sigma). A secondary antibody, PE-labeled goat anti-mouse IgG, was used to produce a signal specific for transfected cells. After PE-labeling, cells were subjected to high speed flow cytometry sorting and the top 5% selected.

[0139] To produce conditioned medium with His-AP-Sp35, the cells that expressed the highest levels of HIS-ApSp35 were selected. The cell lines were adapted to grow in suspension culture in serum free media (BCM16). The titer of His-AP-Sp35 that was produced by these clones was determined by growing cell lines at 37°C for 4-5 passages, then growing the cells to 50% of maximal cell density and culturing them for 10-15 days at 28°C until the viable cell density dropped to 75%. The culture media were harvested, cleared of cells and debris by centrifugation, and the culture supernatants titered for His-AP-Sp35 levels by western blot analysis using anti-human AP antibody (Jackson Labs) as the probe.

[0140] His-AP-Sp35 was purified from the conditioned medium as follows: 400 mL of conditioned medium from CHO cells expressing His-AP-Sp35 was diluted with 400 mL of water. Triethanolamine pH 8.5 was added to 25 mM from a 0.5 M stock and the sample was batch loaded for 2 hours at 4°C onto 6 ml of Fractogel TMAE (EM Industries) anion exchange resin. The resin was collected in a 1.5 cm (I.D.) column, and washed two times with 6 mL of 10 mM HEPES pH7.5, 50 mM NaCl. The AP-Sp35 was eluted from the column with 10 mM HEPES pH 7.5, 200 mM NaCl in 2 mL fractions. Peak fractions were identified by monitoring AP activity and by SDS-PAGE. The flow-through fraction from the TMAE column was further diluted with 300 ml of water and loaded in batch overnight at 4°C onto 6 ml of TMAE resin. The resin was collected and washed as described above and eluted with 10 mM HEPES pH 7.5, 150 mM NaCl. Peak fractions again were identified by monitoring AP activity and by SDS-PAGE. His-AP-Sp35 from the first column was 50% pure and AP-Sp35 from the second column was 90% pure. Under reducing conditions, the His-AP-Sp35 migrated on SDS-PAGE gels with an apparent mass of 130 kDa. While the 90% pure material was appropriate for most investigations, for some studies the His-AP-Sp35 was further purified on Ni-NTA agarose resin (Qiagen). NaCl was added to the elution fractions from the TMAE column to 800 mM and 0.5 M triethanolamine pH. 8.5 and 1M imidazole pH 7.0 was added to 25 mM and 15 mM, respectively. 4.5 ml of the sample was loaded onto a 400 µL NiNTA column. The column was washed three times with 25 mM Triethanolamine pH 8.5, 800 mM NaCl, 15 mM imidazole, and the His-AP-Sp35 was eluted from the column with 200 mM imidazole pH 7.0, 350 mM NaCl, collecting 200 µL fractions. Peak AP-containing fractions were pooled and dialyzed overnight against 250 volumes of 10 mM HEPES pH 7.5, 200 mM NaCl. $MgCl_2$ and $ZnCl_2$ were added to the retentive to 2 and 0.25 mM, respectively. The final product was greater than 95% pure by SDS-PAGE, and ran as a single band with mass of approximately 140kDa under reducing conditions.

[0141] The Sp35 constructs were also engineered as Fc fusions. An Sp-35 LRR-Fc construct was generated by PCR using primers (forward) 5'CTTGACACG GGATCCGCGGCCGCATGCTGGCGGGGGGCGTGAGG3' (SEQ ID NO: 20) and (reverse) 5'GCAGCGGGGCGGGCAGCCCGTGGCCGAGCCTGACAGC

ACTGAGCC3' (SEQ ID NO: 21). The PCR product was inserted into the NotI site of PV90 vector. Sp35 IG-Fc construct was generated by PCR using primers (forward) 5'CTTGACACGGGATCCGCGGCCGCATGCTGGCGGGGGGC GTGAGG3' (SEQ ID NO: 22) and (reverse) 5'GTCCCGGATGCGGGCGCGGG CCGAGCCTGACAGCACTGAGCCCAG3' (SEQ ID NO: 23). The PCR product was inserted into the NotI site of PV90 vector. The proteins were expressed in CHO cells and purified using a protein A sepharose column.

Example 4: Sp35 Binding to NgR1-Expressing Cells

[0142] Four different methods were used to show Sp35 binding to NgR1. First we detected the interaction in a direct binding assay in which the alkaline phosphatase-Sp35 conjugate (AP-Sp35) was incubated with NgR1 expressing cells and binding assessed using a chromogenic AP detection reagent. 90% confluent COS7 cells were grown on 100 mm tissue culture dishes and transfected with NgR1 expressing plasmids using Fugene 6 reagents (Roche). After 48 hours, the transfected cells were washed once with HBH (Hank's balanced salt buffer, 1 mg/ml BSA, 20 mM HEPES, pH 7.0), and then incubated for 1.5 hr at 23°C with 4 μg/ml of the AP-Sp35 fusion protein in HBH. The cells were washed with ice-cold HBH buffer 3 times for 3 min each, then fixed with 3.7% formaldehyde in 20 mM HEPES, pH 7.0, 150 mM NaCl for 15 min, and transferred back into HBH buffer. Endogenous heat-labile AP was heat-inactivated for 2 hours at 67°C. Bound AP-Sp35 was detected by incubation with nitro blue tetrazolium NBT (Roche). Ap-Sp35 bound to COS7 cells expressing human NgR1 receptor, but not to control COS7 cells transfected with the vector alone. A punctate staining pattern for NgR1 was observed, reflecting that only a fraction, probably 50% of the cells, were transfected with the NgR1.

[0143] To better quantify binding, we performed the same experiment but parallel cell samples were treated with 8, 4, 2, 1, 0.5, 0.125, 0.06 μg/ml of the AP-Sp35. Bound AP was incubated with 4-nitrophenyl phosphate and AP activity assessed in a 96-well plate reader (Molecular Devices). From these data, we estimated that the EC50 for binding of Ap-Sp35 to human NgR1 was approximately 6 nM.

[0144] Second, we detected binding of Sp35 to NgR1 in an ELISA approach. ELISA plates (Costar) were coated with 10 μg/ml soluble NgR1-Fc receptors (sNgR310-Fc containing rat NgR1 peptide 35-310 fused to the hinge and Fc of rat IgG1 and sNgR344-Fc containing rat NgR1 peptide 35-344 fused to the rat IgG1) in 0.1M $NaHCO_3$, pH 9.0 for 1 hr at 37°C. The plates were blocked and washed with 25mM Hepes, pH 7.0, 0.1% BSA, 0.1% ovalbumin, 0.1% non-fat dried milk and 0.001% NaN_3. AP-Sp35 protein, 4 μg/ml, was added to the plate and incubated overnight at 4°C. The plates were then washed with 10 mM Tris pH 7.5, 150 mM NaCl and bound AP detected using 10 μg/ml of the chromogenic substrate 4-nitrophenyl phosphate diluted in 0.1 M glycine, 1 mM $MgCl_2$, 1mM $ZnCl_2$ pH 10.5. The OD_{410} was determined in an ELISA reader (Molecular Devices) equipped with the Softmax program. AP-Sp35 bound to immobilized sNgR-344-Fc but not the sNgR-310-Fc protein, indicating that the longer version of NgR1 was required for Sp35 binding. We were able to compete the binding of AP-Sp35 to sNgR344-Fc NgR1 by 80% by pre-incubating the AP Sp35 with 100-fold excess of sNgR344-Fc. No competition of binding was seen using a hedgehog-rat Ig1 fusion protein as a rat Ig fusion control protein.

[0145] Third, we detected the binding of Sp35 to NgR1 by co-immunoprecipitating Sp35 with NgR1. For this study, 80% confluent COS7 cells, grown on 100 mm tissue culture dishes, were transfected with plasmids encoding Sp35-hemagglutinin (Sp35-HA) and NgR-FLAG using Fugene 6 reagents (Roche) 48 hours after transfection, cells were harvested and lysed in 1ml lysis buffer (50 mM HEPES, pH 7.5, 150 mM NaCl, 1.5 mM $MgCl_2$, 1 mM EGTA, 1% Triton X-100 and 10% glycerol) at 4°C for 30 min. The lysate was then centrifuged at 14,000 x g for 15 min, and the supernatants collected and incubated 4°C overnight with agitation, using anti-HA affinity matrix (Roche). The samples were washed 3 times with 1 ml of lysis buffer, then boiled for 3 minutes in Laemmli sample buffer, subjected to 4-15% SDS-PAGE, and analyzed by immuno-blotting with Anti- FLAG M2 antibody (Sigma). The anti-HA tag affinity resin collected a complex containing both Sp35-HA and FLAG-NgR, as is evident by the presence of FLAG. This complex was not seen in lysates from control transfections in which cells were treated with Sp35-HA plasmid or FLAG-NgR1 plasmid alone, or

cells that were co-transfected with the Flag-NgR1 and a HA-tagged control protein that does not bind NgR1.

[0146] Sp35-HA was made as follows. The Sp35 signal sequence and extracellular domain (amino acids 1-531) was PCR amplified using primers 5'ATATTCTAGAATGCTGGCGGGGGGCGTGAG3' (SEQ ID NO: 24) and 5' ATATACTAGTGTCGTTGCCGCCCGCGTTGG3' (SEQ ID NO: 25) containing XbaI and SpeI sites (underlined). The PCR product was digested by XbaI and SpeI and inserted into the vector pCGCHA between the Xba I and Spe I sites. The sequence of the insert was confirmed by DNA sequencing. The FLAG NgR1 construct was gift from Dr. Zhigang He (Nature, Vol 420, Nov 7, 2002).

[0147] Fourth, we showed that Ap-Sp35 bound to rat cerebellum granular neurons (CGN) which express NgR1. For this experiment, 90% confluent postnatal day 8 CGN cells were grown on 100 mm tissue culture dishes. After 48 hours, the cells were washed once with HBH buffer, and then incubated with 4 µg/ml of AP-Sp35 in HBH buffer for 1.5 hour at 23°C. The cells were then washed with ice-cold HBH 3 times for 3 minutes each, then fixed with 3.7% formaldehyde in 20 mM HEPES, pH 7.0, and 150 mM NaCl for 15 min, and transferred back to HBH. Endogenous heat labile AP was heat-inactivated for 2 hours at 67°C. Bound AP-Sp35 was detected by incubation with nitro blue tetrazolium NBT (Roche). AP-Sp35 bound to postnatal day 8 cerebellum granular neurons, which express NgR1. The binding of AP-Sp35 to the neurons was inhibited by treating the CGN with PIPLC (5 units/ml) which cleaves most GPI anchored proteins from membrane surfaces. Since NgR1 is a GPI-linked protein, this result further supports the notion that Sp35 is binding to NgR1 on CGN cells.

Example 5: Co-localization of Sp35 with NgR1

[0148] To determine whether Sp35 and NgR1 are expressed in the same neurons, we performed a co-localization study. 4% paraformaldehyde-fixed rat p8 primary granular neuron cultures were incubated with antibodies against Sp35 and NgR1 (Santa Cruz), and then with the appropriate Alexa-labeled secondary antibodies (Molecular Probes Inc.). The cells were visualized by con-focal fluorescence microscopy. Neurons were intensely stained by the Sp35 and NgR1 antibodies.

Both proteins were expressed in the cell bodies and axons of neurons. To aid in the co-localization analysis, different colored probes were used for the 2 types of antibodies. When the stains (red for NgR positive cells and green for Sp35 positive cells) were merged we saw a yellow color throughout the cell indicating that the two proteins were co-localized within the neurons.

Example 6: NgR1 Binding Sites within Sp35

[0149] We used deletion mapping to define the specific domains of Sp35 involved in NgR1 interactions. The following deletion constructs were made using the Stratagene Quikchange Mutagenesis kit. We verified all vector constructs by DNA sequencing of the modified inserts.

[0150] His-AP-Sp35b which contains the leucine rich repeat domain of Sp35 plus the basic region (a.a 34-432) was cloned from the His-AP-Sp35 (a.a. 34-532) vector by PCR. Primers used were 5'CCCAGCAGGTGTTTGTGGACGAGTG ATCTAGGGCCGCGGATCCCTG-3' (SEQ ID NO: 26) and 5'-CAGGGATCCG CGGCCCTAG ATCACTCGTCCACAAACACCTGCTGGG-3' (SEQ ID NO: 27).

[0151] His-AP-Sp35d, which encodes the Ig domain of Sp35 plus the basic region (a.a 417-531), was cloned from the His-AP- Sp35a (a.a. 37-531) vector by PCR. Primers used were 5'CGCCGCGCACCCGGGTGAATTCCGCGCCCGC ATCCGGGACCGC-3' (SEQ ID NO: 28) and 5'-GCGGTCCCGGATGCGGGCGC GGAATTCACCCGGGTGCGCGGCG-3' (SEQ ID NO: 29).

[0152] His-AP-Sp35e, which encodes only the Ig domain (a.a 425-531), was cloned from the His-AP- Sp35(aa 34-532) vector by PCR. Primers used were 5'-CGCCGCGCACCCGGGTGAATTCGCCCAGCAGGTGTTTGTGGAC-3' (SEQ ID NO: 30) and 5'-GTCCACAAACACCTGCTGGGCGAATTCACCCG GGTGCGCGGCG-3' (SEQ ID NO: 31).

[0153] A commercial mutagenesis kit and protocol (Stratagene Quikchange) were used to mutate amino acid 456 (from arginine to glutamic acid) and amino acid 458 (from histidine to valine) of Sp35 in the vector His-AP-Sp35 (34-532). The primers used were 5'-CATCCTCTGGCTCTCACCCGAAAAGGTACTGG TCTCAGCCAAGAGC-3' (SEQ ID NO: 32) and 5'-GCTCTTGGCTGAGACCA GTACCTTTTCGGGTGAGAGCCAGA GGATG-3' (SEQ ID NO: 33).

[0154] His-AP-Sp35 deletion constructs (FIG. 3) were engineered in pV90 expression vectors and expressed in 293 cells. The conditioned medium was collected and the AP adducts purified by sequential chromatography steps on Fractogel TMAE resin and NiNTA agarose. The purified proteins were tested for binding to NgR1 expressed on COS7 cells. The three constructs all bound weakly to Sp35. These results indicated that the Sp35 LRR repeat 1-14 (amino acid 34 to 417) and the Ig domain of Sp35 (amino acid 425-531) both contribute to Sp35 binding to NgR1. The Ig domain showed higher affinity than the LRR domain.

[0155] A structural model for the Ig domain of Sp35 was generated using the NCAM crystal structure as a framework (Rasmussen et al., 2000, Nat. Struct. Biol. 7:389-393). From this model, we observed a loop (residue numbers 454-458, amino acids: SPRKH; SEQ ID NO: 34) that might be involved in binding. To test this hypothesis, we engineered an Sp35 construct in which residues R at 456 and H at 458 were changed into E and V, respectively. When this construct was tested for NgR1 binding, we observed a >10 fold drop in signal. As an alternative approach to test the contribution of this loop region in binding, we synthesized a peptide corresponding to the sequence LSPRKH (SEQ ID NO: 10) that we cyclized by adding cysteines at the N and C terminus of the peptide. Upon binding to NgR1, this peptide blocks, inhibits, or interferes with the function of NgR1.

Example 7: Sp35 induces p8 CGN fasciculation

[0156] To determine the biological function of Sp35 in neurons, we incubated Sp35-Fc with postnatal day 8 granular neurons to see if Sp35 can regulate neurite out-growth. Labtek culture slides (8 well) were coated with 0.1 mg/ml poly-D-lysine (Sigma) before spotting of Sp35-Fc protein (16 μg/well protein). The slides were dried overnight, then rinsed and coated with 10 μg/ml laminin (Gibco). Cerebellum granular neurons from postnatal day 8 were dissociated and seeded onto the precoated slides. The slide cultures were incubated at 37°C in 5% CO_2 for 24 hours. The slides were then fixed in 4% paraformaldehyde containing 20% sucrose and stained with anti ßIII tubulin (Covance TUJ1). After 24 hours, the CGN showed clear fasciculation morphology as evident by bundling of the neurons. The fasciculation was not seen in the untreated cells or Fc protein-coated sample controls.

Example 8: Effects of Sp35 on RhoA activation/inactivation:

[0157] Sp35-Fc induced postnatal cerebellum granular neurons to undergo fasciculation. Because the signaling molecule RhoA is known to be involved in fasciculation, we determined if Sp35-Fc can regulate RhoA functions in neurons. We performed the RhoA activation experiment as follows: 293 cells or COS7 cells were transfected with expression vectors containing combinations of RhoA, Sp35 or NgR1 using Fugene 6 reagents (Roche). 48 hr post-transfection, cells were serum starved overnight, then lysed in 50 mM Tris, pH 7.5, 1% Triton X-100, 0.5% sodium deoxycholate, 0.1% SDS, 500 mM NaCl, 10 mM $MgCl_2$, plus a protease inhibitor cocktail. Cell lysates were clarified by centrifugation at 13,000xg at 4°C for 5 minutes, and 95% of the supernatants were incubated with 20 µg of an immobilized GST-Rho binding domain affinity matrix (Rhotekin beads, Upstate Biotechnology) at 4°C for 45 minutes. The beads were washed 3 times with wash buffer (50 mM Tris, pH 7.5, 1% Triton X-100, 150 mM NaCl, 10 mM $MgCl_2$, with protease inhibitors). GTP-bound Rho was eluted from the beads by heating at 95°C for 5 min in SDS-PAGE sample buffer. Bound and total Rho proteins were detected by western blotting using a monoclonal antibody against RhoA (Santa Cruz). COS7 and HEK293 cells transfected with Sp35 induced RhoA activation, as evident by an increase in the amount of RhoA-GTP detected in the blot following transfection with the Sp35 gene. A further enhancement in RhoA-GTP was observed following treatment with Sp35-Fc. In contrast to the increase in RhoA-GTP following transfection with Sp35 alone, when cells were transfected with Sp35 and NgR1, RhoA was partially inactivated. Treatment of these cells with Sp35-Fc resulted in further inactivation of RhoA.

[0158] We confirmed a signaling response by Sp35 using a FLIPR assay (Molecular Devices) to determine the affects of Sp35 treatment on Ca^{++} flux. We observed a significant Ca^{++} flux in cells expressing Sp35 with treatment of Sp35-Fc, but not in the control cells treated with Sp35-Fc. The Ca^{++} flux was reduced when cells that had been co-transfected with NgR1 and Sp35 were treated with Sp35 –Fc fusion protein.

Example 9: Sp35 Protein Interaction with Itself

[0159] Because LRR domains frequently are involved in homotypic interactions, and we observed that the addition of soluble Sp35 to Sp35-transfected cells caused an increase in RhoA-GTP over what was observed with Sp35 transfections alone, we tested for Sp35 binding to itself. To perform this test, we used co-immunoprecipitation. COS7 cells 80% confluent, grown on 100 mm tissue culture dishes, were transfected with plasmids Sp35 HA or Sp35-FLAG, or both, using Fugene 6 reagents (Roche). Forty-eight hours after transfection, cells were harvested and lysed in 1ml lysis buffer (50 mM HEPES, pH 7.5, 150 mM NaCl, 1.5 mM $MgCl_2$, 1 mM EGTA, 1% Triton X-100 and 10% glycerol) at 4°C for 30 min. The lysate was then centrifuged at 14,000xg for 15 min, and the supernatants collected and incubated, at 4°C overnight with agitation, with an anti-HA affinity matrix (Roche). The samples were then washed 3 times with 1 ml of lysis buffer, boiled in Laemmli sample buffer, subjected to 4-15% SDS-PAGE, and analyzed by immuno-blotting with anti-FLAG antibodies. The Anti-HA antibody resin captured a complex that contained Sp35-FLAG, as determined by Western blotting. This indicated a direct interaction of Sp35 with itself. We also treated cells transfected with HA-Sp35 with Sp35-Fc and used a similar immunoprecipitation approach to show that HA-Sp35 bound to Sp35-Fc.

[0160] Sp35-FLAG was made as follows. Sp35 gene extracellular domain (a.a. 1-531) was PCR amplified using primers 5'AATTAAGCGGCCGCATGCTGGCG GGGGGCGT3' (SEQ ID NO: 35) and 5'AATTAAGCGGCCGCTTTGTCATGT'3 (SEQ ID NO: 36) containing NotI sites (underlined). The PCR product was digested by NotI and inserted into the NotI site of vector pV90. The DNA sequence of the insert was confirmed by DNA sequencing.

Example 10: *In Vivo* Transplantation of Sp35-Transformed Cells

[0161] To determine the biological function of Sp35 in spinal cord injured rats, we infected cortical primary cultured cells (mixed cultures) with retrovirus expressing full length Sp35 or a retrovirus control, for delivery into the injured epicenter of rat spinal cords. $2x10^6$ cells were introduced, and the rats were sacrificed at day 10. The spinal cords were fixed in 4% paraformaldehyde

overnight, then dehydrated in 70% , followed by 95% ETOH. Tissue samples were imbedded in paraffin. Sections (10 microns thick) were used for immunohistochemical staining. Rats that received Sp35-expressing cells, in comparison to control, show less axon retraction and more -ßIII tubulin staining near the epicenter. Increased neuronal survival in the injured rats receiving Sp35 was observed.

[0162] The Sp35 retrovirus construct was made as follows: The Sp35 gene was PCR amplified using primers 5'-GATTACTCGAGATGCTGGCGGGGGGCGT GAGG-3' (SEQ ID NO: 37), containing an XhoI site (underlined), and 5'CGCGGGAATTCTCATATCATCTTCATGTTGAACTTG-3' (SEQ ID NO: 38), containing an EcoRI site (underlined). The PCR product was digested with XhoI and EcoRI , then ligated into the Retrovirus vector pMIG (which contains IRES-GFP), which was previously cleaved with XhoI and EcoRI. The new vector was named pMMC078. All isolates of pMMC078 contained inadvertent point mutations, so two isolates of pMMC078 were ligated together. pMMC078.6 was cut with XhoI and AccI and pMMC078.7 was cut with XhoI and AccI. These two fragments were ligated together to make the final correct plasmid, pMMC089. The DNA sequence of the insert was confirmed by DNA sequencing. Sp35 retrovirus was made as described. 293G cells were split the day before transfection. 8 μg Sp35-retrovirus DNA was used to transfect $5x10^6$ cells by lipofectamine (Invitrogen). The condition medium was harvested after 92 hours post-transfection. The conditioned medium was centrifuged at 5000g for 10 minutes, and the supernatant used as a Sp35 retrovirus stock. This stock was stored at 4°C for 1 week or –80°C for 6 months.

Example 11: Animal Model of Spinal Cord Injury

[0163] All surgical procedures are performed using aseptic technique. For 1 week prior to any surgical manipulation animals are handled. Ampicillin 100mg/kg SC is administered prophylactically prior to and after surgery to reduce the incidence of bladder infection injury.

[0164] Animals are anesthetized using Midazolam at 2.5 mg/kg IP in conjunction with Isoflurane 2-3% in O_2 to deep anesthesia as measured by toe pinch. Animals

are maintained on a circulating water heating pad for the duration of surgery and recovery. Ocular lubricant are used to prevent corneal drying and Atropine 0.05mg/kg SC are given to reduce excess salivation. A small incision is made in the skin and the muscle retracted to expose the vertebrae. A dorsal laminectomy at the spinal level L6 (and L7 if placement of an intrathecal catheter is necessary, see below) is performed, L6/L7and the adjacent spinous processes rigidly fixed in a spinal frame (David Kopf Instruments). A dorsal hemisection is performed at L6 with fine iridectomy scissors completely interrupting the main dorsomedial and the minor dorsolateral coticospinal tract (CST) components. Following surgery, the laminectomy site is covered with a protective material such as Durafilm, and the overlying muscle sutured with 4.0 chromic gut to protect the exposed spinal column. The skin is sutured and wiped with betadine solution.

[0165] Functional recovery of animals IS evaluated using the Basso Beattie and Bresnehan (BBB) scoring method commonly used to evaluate rats after spinal cord injury. This method quantifies the hind limb function of rats by detailed analysis of joint movement and weight bearing ability. Rats are evaluated the day after spinal cord injury then weekly thereafter.

[0166] Immediately after CST transection, adenovirus expressing Sp35 or GFP or control virus (10^{10} particles) are injected at the site of transection and regions immediately caudal and rostral to the injury site. A total of 10µl of Adv are injected at 5 different sites (4ul/site). For the intrathecal administration of Sp35 protein, a small hole is made into the dura of the spinal cord at 2mm caudal to the lesion L7 and an intrathecal catheter is inserted into the subarachnoid space at L7. The catheter is slowly and gently slid above the spinal cord about 1mm caudal to the lesion. The portion of the catheter lying outside the intrathecal space is firmly sutured in place to the surrounding tissue. A primed mini-osmotic pump (Alza corp.) containing the test material (Sp35 protein or control protein) is connected to the exposed end of the guide cannula and inserted into the subcutaneous space. Following surgery, the laminectomy sites are covered with a protective material such as Durafilm and the overlying muscle sutured with 4.0 chromic gut to protect the exposed spinal column. The skin is sutured and wiped with betadine solution.

[0167] Histological Analysis: Tract tracing surgery occurs at the time of the surgery to induce spinal cord injury. The skin on the head is shaved and wiped with Betadine and 70% alcohol. The animal is placed in a stereotaxic frame. The scalp is incised longitudinally and the periosteum is scraped from the calvaria. A hole is drilled in the skull approximately 1-2 mm in diameter, and a glass microliter needle is inserted vertically into 8 locations in the motor cortex (coordinates are determined according to the rat brain atlas of Paxinos and Waston, 1997). Approximately 5µl of tract tracer material (e.g., Biotin dextran amine , 10,000M.Wt) is injected and the needle is left in place for an additional five minutes to allow diffusion of the solution. After needle removal, the hole in the scull cap is plugged with gel foam and the scalp stapled closed over the injury site. Animals are allowed to recover and receive post-operative care (described below). Four to ten weeks later the animals are deeply anesthetized (Inactin 100-110mg/kg ip) and perfused for histology as described below. The tract tracer is carried by anterograde transport mechanisms down the cortical spinal tract towards the caudal end of the spinal cord and provides a means to quantify anatomical connectivity within the corticospinal tract.

[0168] For immunohistochemistry experiments, animals are deeply anesthetized with Inactin (100-110 mg/kg IP) 2-8 weeks after surgery to induce injury. The chest cavity is opened and the heart is exposed, to allow for perfusion. A cannula is inserted into the left ventricle through which 100cc ice-cold PBS is pushed slowly (a hole will be cut in the right ventricle to allow fluid escape). This is followed by a slow but steady drip of 4% paraformaldehyde (50-100ml) until fixation of eyes/ears/toes is obvious. Spinal cords are removed, with care to minimize alteration of the injury site, frozen in OCT, sectioned, and processed for immunohistochemistry. Other tissues optionally are also collected for later analysis. The animals receiving adenovirus Sp35 showed increased axon sprouting as determined by ßIII tubulin staining for neuron axon.

Example 12: Sp-35 Viral Vector Constructs

[0169] A pMIG-derived Sp-35 viral vector was made as follows. The full-length Sp35 coding sequence was PCR amplified using primers 5'-GATTACTCGAGA TGCTGGCGGGGGGCGTGAGG-3' (SEQ ID NO: 37), containing an XhoI site,

and 5' CGCGGGAATTCTCATATCATCTTCATGTTGAACTTG-3' (SEQ ID NO: 38), containing an EcoRI site. The PCR product was cut with XhoI and EcoRI , then ligated into the Retrovirus vector pMIG (Cheng et al, 1996, Nat. Biotechnol.145:576) which was cut with XhoI and EcoRI. This vector was designated pMMC078. All isolates of pMMC078 contained point mutations, so two isolates of pMMC078 were ligated together. Vector pMMC078.6 was cut with XhoI and AccI and pMMC078.7 was cut with XhoI and AccI. These two fragments were ligated to make the plasmid, pMMC089.

[0170] A pMIG-derived Sp35-HA viral vector was made as follows. A fragment encoding Sp35 amino acids 326-614 in frame with the HA sequence was obtained by using PCR with primers 5'- GCCTTCCGCGGCCTCAACTACCTGCGCGTG CTC-3' (SEQ ID NO: 39), containing a SacII site, and 5'-CCGGAATTCTCA *AGCGTAATCAGGAACGTCGTAAGGGTA*TATCATCTTCATGTTGAACTTGCG GGGCGCGTCGGC-3' (SEQ ID NO: 40), with pMMC089 serving as a template. The longer primer includes the HA coding sequence (italics) after Sp35 codon 614 and before the EcoR I site. The PCR product was then cut with Sac II and EcoR I, and used to replace the Sac II-EcoR I fragment containing wild-type Sp35 codons 326-614 in the pMIG-derived retroviral vector.

[0171] An Sp35-baculovirus HA vector was made as follows. The Sp35-HA coding sequence from Sp35-HA retroviral vector was cut out with Xho I and EcoR I, blunt-ended, and cloned into Bgl2-fill in site of baculo viral shuttle vector pBV-CZPG (U.S. Pat. Nos. 6,190,887; and 6,338,953), replacing LacZ gene under CMV promoter.

[0172] An Sp35-adenoviral vector was made as follows. Sp35-IRES-GFP coding sequence from the Sp35-retroviral was cut out with Xho I-fill in and Nhe I, then cloned into the EcoR1-fill in/Nhe I sites of the Adenovirus shuttle vector pDC315, under minimal CMV promoter.

Example 13: Animal Model of Remyelination

[0173] Female Long Evans rats are used in all studies. Rats are anaesthetized using Isoflurane and the T3/4 exposed and a dorsal hemi-laminectomy performed. The chemical demyelinating agent, lysolecithin (3µl of 1% lysolecithin in 0.9% saline), is then injected into the right side of dorsal columns of the spinal cord 0.5-

1mm below the surface of the cord). Appropriate analgesic treatment is administered before and after surgery.

[0174] Three days later, the injection site is re-exposed (under isoflurane anesthesia, with appropriate analgesic treatment) and the following therapies injected into the injured spinal cord and an adenovirus vector encoding protein Sp35/control protein is injected into the injury site. 10^{10} particles of adenovirus encoding Sp35 or GFP control in a volume of 10 μl will be injected into injured rat spinal cord in up to 5 different sites in and around the site of lysolecithin-induced demyelination. A volume of not greater than 2 μl is injected at each of the 5 injection sites. For histological analysis of spinal cord demyelination/remyelination 2, 3, 4 or 6 weeks after surgery, animals are deeply anesthetized with inactin (100-110 mg/kg ip) and perfused with fixative via the heart. The spinal cord is then removed and processed for analysis. The animal receiving Sp35 treatment showed increased axon myelination as determined by IHC using anti-MBP protein antibody or luxol fast blue.

Example 14: Sp35 RNAi

[0175] To address the role of Sp35 in brain function, we introduced the lentivirus Sp35 RNAi into postnatal 8 CGN cells. Sp35 RNAi infected cells had shorter neurites and higher rates of proliferation than control cells. These results indicate a role for Sp35 in regulating RhoA activation.

[0176] Murine and rat Sp35 DNA sequences were compared to find homologous regions to use for candidate shRNAs. CH324 was constructed by annealing oligonucleotides LV1-035 and LV1-036 and ligating to Hpa1 and Xho1 digested pLL3.7. The oligonucleotides were purchased from MWG. The sequences are:

LV1-035 (sense oligo) 5'TGATCGTCATCCTGCTAGACTTCAAGAGAGTCT AGCAGGATGACGATCTTTTTTC (SEQ ID NO: 41)

LV1-036 (antisense oligo)

5'TCGAGAAAAAAGATCGTCATCCTGCTAGACTCTCTTGAAGTCTAGCAG GATGACGATCA (SEQ ID NO: 42).

[0177] Prior to producing virus, DNA from pLL3.7 or candidate shRNA in pLL3.7 were cotransfected with murine SP35-HA tagged plasmid at a ratio of 5 to 1 into CHO cells in 6 well format. Knockdown was analysed by western blot

detection of SP35-HA tag from transfected CHO cell lysates as well as by northern blot of total RNA prepared from duplicate wells. The blot was probed with a 0.7 kb fragment of mSP35. Assays were performed 48 hours post-transfection (data not shown). Viruses were produced from the best candidate for use in rat neuronal cultures. The vector, additional methodology and virus production were as described in Rubinson et al. "A lentivirus-based system to functionally silence genes in primary mammalian cells, stem cells and transgenic mice by RNA interference." Nat. Genet. 33, 401-6 (2003).

Example 15: RhoA Activation

[0178] COS7 cells co-expressing NgR1 and SP35 showed no changes in RhoA/GTP levels in response to OMgp. This suggested that the SP35/NgR1 complex is not sufficient for mediating signal transduction by a myelin inhibitor.

[0179] We explored the possibility that a ternary complex of SP35/NgR1/p75 mediates signaling. Two approaches were used to evaluate interactions between SP35, NgR1 and p75. First, binding was evaluated in a direct binding assay using an AP-SP35 conjugate. The AP-SP35 conjugate bound weakly to p75-expressing cells. AP-P75 bound to NgR1 expressing cells. The binding of AP-SP35 to NgR1 and p75 were measured by ELISA (FIG. 4). Second, binding of SP35 to NgR1and p75 was evaluated by a co-immunoprecipitation from COS7 cells co-expressing SP35 NgR1 and p75. An anti-NgR1 antibody immunoprecipitated a complex containing SP35 and p75. An anti SP35 antibody also immunoprecipitated a complex containing p75. The interaction and co-immunopreicipitation data provided evidence for a direct interaction between SP35, NgR1and p75. We used confocal microscopy and antibodies against SP35, p75 and NgR1 to show that SP35, NgR1 and p75 co-localize to cell bodies and axons of p7 CG neurons from rat.

[0180] Next we showed that the combination of SP35, NgR1 and p75 is sufficient for the activities of the myelin inhibitor. Non-neuronal COS7 cells were engineered to express all three components. Using these cells, we showed that RhoA/GTP levels were up-regulated by OMgp. OMgp-Fc treatment increased RhoA/GTP levels in SP35/p75/NgR1 co-expressing cells, as compared to other combinations of these three components. We confirmed expression of the proteins

by Western blotting of COS7 cells lysates. The affinity of myelin inhibitors binding to NgR1 were not affected by the presence of p75 or p75 and SP35. The combined results support a model whereby a ternary complex of NgR1, SP35 and p75 is required for RhoA regulation in the presence of NgR1 ligands (FIG. 5).

[0181] SP35 contains a cytoplasmic domain that has potential direct or indirect involvement in signaling. To determine the role of the cytoplasmic domain, we produced a cytoplasmic domain truncation of SP35 (amino acids 34 to 576 of SEQ ID NO: 2), to function in a dominant negative manner by forming an unproductive, ternary complex incapable of signaling. We designated this molecule with the cytoplasmic domain truncation "DN-SP35" (for dominant negative SP35). We transfected postnatal day 7(p7) CG neurons with full-length SP35 or DN-SP35, and then assayed for response to the inhibitory myelin components (Omgp, myelin and Nogo66). As shown in FIG. 6, DN-SP35 transfected cells failed to respond to the inhibitory myelin components and showed longer neurites than controls. In contrast, cells transfected with the full length SP35 construct showed enhanced response to the inhibitory substrates, and had shorter neurites as compared to controls. This demonstrated that DN-SP35 acts as a competitor to attenuate neurite outgrowth inhibition caused by myelin components. We expected that exogenous, soluble SP35-Fc also would bind NgR1 and block the action of inhibitory substrates. As shown in FIG. 7, SP35-Fc reduced the neurite outgrowth inhibition by Omgp, Nogo66 and MAG.

Example 16: Neuroprotective Activity

[0182] Equal numbers of rat p6 cerebellar granule neurons was plated in each well of a 12-well cell culture plate in the presence or absence of 50 nM of sp35-Fc protein. These poly-D-lysine plates have been pre-coated [dried down] with 10 μg of CNS myelin, or 200ng of Nogo66, MAG and OMgp or control-Fc. The neuronal cultures were maintained for 1-7 days at 37°C and 5%CO_2. The neurons were healthy and grew well in the PBS control wells independent of sp35-Fc treatment with full neurite extension [determined by neuronal specific marker, βIII tubulin] as examined after 3 days. In the absence of sp35-Fc, the neurons did not grow well in the wells coated with myelin, Nogo66, MAG and OMgp. There was minimum neurite sprouting [short and distorted] and the neurons did not

appear to be healthy, with rounded cell body and condensed nuclear materials. DAPI staining demonstrated that the number of neurons detected in these wells was less than that in the PBS control wells, suggesting neuronal loss. In the presence of sp35-Fc, long neurites were present and the neurons appeared healthy. DAPI staining demonstrated a higher neuronal number in these wells than those that did not receive the sp35-Fc. The data are summarized in Table 2 below.

Table 2

In the absence of sp35-Fc

dried down substrate:	OMgp/Nogo/MAG/ myelin	Fc control
neurite extension	short distorted	long extended
cell body morphology	rounded	spread
nuclear materials	condensed	clear
neuron number at the end of expt	reduced	same as control Fc

In the presence of sp35-Fc

dried down substrate:	OMgp/Nogo/MAG/ myelin	Fc control
neurite extension	long extended	long extended
cell body morphology	spread	spread
nuclear materials	clear	clear
neuron number at the end of expt	less reduced that control FC	same as control Fc

These data indicated that a soluble form of Sp35, e.g., Sp35-Fc, possesses neuroprotective activity.

[0183] In spinal cord hemi-transected (T9, SCT) rats, β-III tubulin staining of the spinal cord sections showed a substantial loss of neurons at the lesion site. A recombinant virus expressing sp35 was used to infect the SCT animals at the lesion site. Histological staining of these spinal cords showed an increased number of neurons around the lesion site compared to the control group that was infected with the vector virus. This is consistent with the *in vitro* experimental findings

described above, and further indicates neuroprotective properties associated with Sp35.

Example 17: Sp35 in Animal Model of Spinal Cord Injury

[0184] Since Sp35-Fc reduced neurite outgrowth inhibition caused by OMgp, Nogo-66 and MAG *in vitro*, we expected the molecule to promote functional recovery of CNS injuries *in vivo*. To confirm this, we administered Sp35-Fc to spinal chord hemisected rats, i.e., an animal model of acute CNS trauma. As shown in FIG. 8 and FIG. 9, Sp35-Fc treated rats demonstrated significantly improved functional recovery, compared to control rats treated with IgG.

[0185] Spinal cord injury and behavioral analysis were performed as follows. All surgical procedures were performed in accordance with the guidelines of the Biogen Institutional Animal Use and Care Committee. Female Long Evans rats (190-210g, Charles River, Wilmington, MA) were anesthetized using 2.5mg/kg Midazolam, I.P. and 2-3% Fluothane in O_2. A dorsal laminectomy was performed at spinal level T6 and T7. A dorsal hemisection was performed, completely interrupting the main dorsomedial and the minor dorsolateral corticospinal tract (CST) components. Immediately after CST transection an intrathecal catheter was inserted into the subarachnoid space at T7 and connected to a primed mini-osmotic pump (Alzet model 2004) inserted into the subcutaneous space. Mini-osmotic pumps delivered Hu IgG isotype control protein (5mg/ml, n=5, Pharmingen), PBS (n=3) soluble Hu Sp35-Ig fusion protein (4.3mg/ml, n=8) at a rate of 0.25 µl/h. Following surgery, the laminectomy site was sutured and the skin wound stapled closed. Postoperative care included analgesia (Buprenorphine 0.05mg/kg s.c.) for 3 days and antibiotic treatment (Ampicillin 100mg/kg s.c.twice daily) for 7 days after surgery. Bladders were expressed manually twice a day for the duration of the study (28 days) or until return of function (the time of which was noted). All animals were blindly scored using the open-field BBB scoring system (Basso et al., 1995, *J. Neurotrauma* 12:1-21; Ono et al., 2003, *J. Neurosci.* 23:5887-5896). Rats were evaluated the day after CST transection (day 2) and weekly thereafter for 4 weeks using the Basso-Beattie-Bresnahan (BBB) locomotor rating scale. Investigators were blinded to the treatment groups for the duration of the study.

Example 18: Neuronal survival and axon regeneration in the rubro-spinal tract (RST) hemi-section injury model:

[0186] We also investigated the effects of Sp35 treatment on the regeneration of neurons in the rubro-spinal tract which directly contribute to locomotion.

[0187] Adult 9-week-old Sprague-Dawley rats (200-250 g) were anesthetized with an intraperitoneal injection of ketamine (80 mg/kg) and xylazine (8 mg/kg). Under an operating microscope, a dorsal laminectomy was performed and the seventh thoracic spinal vertebra (C7) identified. After opening the dura mater, a right hemi-section was performed at spinal cord level C7 using a pair of spring scissors. Following spinal cord hemi-section, animals received a piece of gelfoam soaked with either 10μl of a 2μg/ml solution of Sp35-Fc, or 10μl of a 2μg/ml solution of human Ig, or 10μl PBS, placed on top of the lesion site. After the operations, animals in each group were subdivided for axonal tracing and behavioral analysis. The animals for axonal tracing (n = 5 for each group) and behavioral analysis (n = 7 for each group) were allowed to survive for 1 month.

[0188] Fluoro-Gold (FG, 6% w/v, Fluorochrome) was used to label the RST neurons that had regenerated their axons across the injury scar and reentered the caudal spinal cord. Two days prior to the end of the post-injury survival period (1 month), animals were anesthetized with an intraperitoneal injection of ketamine (80 mg/kg) and xylazine (8 mg/kg). A dorsal laminectomy was carried out and the T2 spinal segment was identified. FG at a volume of 0.5 ml was manually injected into the right T2 spinal cord using a Hamilton syringe. Two days later, the animals were anesthetized and sacrificed with a lethal dose of ketamine (150 mg/kg) and xylazine (8 mg/kg) and they were perfused intracardinally with normal saline, followed by 400 ml of fixative containing 4% paraformaldehyde in 0.1 x PBS. The brains and spinal cords were removed, postfixed with paraformaldehyde overnight, and then placed in 30% phosphate-buffered sucrose. Brain and spinal cord tissue were cut into 30 mm sections on a cryostat and mounted onto gelatin-coated slides. The number of FG-labeled RST neurons on the lesion side was expressed as a percentage of the total number of FG-labeled neurons on the contra-lateral intact side. This percentage among groups was compared statistically using one-way ANOVA followed by a Tukey-Kramer multiple comparisons test. As

shown in Table 3, Sp35-Fc at 2 μg/ml promoted the survival of rubro-spinal tract (RST) neurons.

Table 3

Treatment	Percent Survival of RST Neurons (± S.E.M.)
PBS	17.1 ± 2
Sp35-Fc	31.9 ± 1.5
Control-Fc	14.5 ± 2.1

[0189] For behavioral analysis, the use of forelimbs during spontaneous vertical exploration was examined 1 month after different treatments as described (Liu et al., 1999) with minor modifications. Rats were placed in a clear Plexiglas cylinder (15 cm in diameter and 30 cm high) that encourages use of the forelimbs for vertical exploration for 5 min. The following behaviors were scored: (1) independent use of the left (unimpaired) or right (impaired) forelimbs for contacting the wall of the cylinder; and (2) simultaneous use of both forelimbs to contact the wall of the cylinder. The vertical exploration behavior was expressed in terms of (1) percentage use of left (unimpaired) forelimb relative to the total number of impaired, unimpaired, and both limb use; (2) percentage use of right (impaired) forelimb relative to the total number of impaired, unimpaired, and both limb use; and (3) percentage use of both forelimbs relative to the total number of impaired, unimpaired, and both limb use. The differences between groups were tested by one-way ANOVA followed by Bonferroni post hoc analysis. Sp35-Fc treated animals showed significantly improved front limb movement: 30% usage for both forelimbs in Sp35-1-Fc treated animals versus 10% usage for both forelimbs in control-Fc or PBS-treated animals; 55% left (unimpaired) limb usage versus 80% usage in control-Fc or PBS-treated animals; and 29% right (impaired) limb usage versus approximately 15% in control-Fc or PBS treated animals.

Example 19: Sp35-Fc promotes retinal ganglion cell (RGC) survival in the optic nerve transection model

[0190] We further confirmed the activity of Sp35 using the optic nerve transection model, which investigates factors that affect neuronal function. Young adult female Sprague Dawley (SD) rats were used in this study. The right optic nerve of each animal was transected intraorbitally 1.5 mm from the optic disc. A

piece of gelfoam soaked with 6% Fluoro-Gold (FG) was applied to the newly transected site right behind the optic disc to label the surviving retinal ganglion cells (RGCs). The animals were divided into 6 groups (n=6 in each group) receiving either Sp35-Fc, human IgG1, or just PBS, by intravitreal injection. The volume of each intravitreal injection was 4 ml while the dosage of each injection was 2 mg. The intravitreal injections were performed immediately after the optic nerve transection.

[0191] All animals were allowed to survive for 1 week. Two days before sacrificing the animals, the left optic nerve of each animal was transected and 6% FG were used to label the surviving RGCs to serve as the internal control. Animals were sacrificed with an overdose of Nembutal and the retinas dissected in 4% paraformaldehyde. Four radial cuts were made to divide the retinas into four quadrants (superior, inferior, nasal and temporal). The retinas were then post-fixed in the same fixative for 1 hour before they were flat-mounted with the mounting medium (Dako). The slides were examined under a fluorescence microscope using an ultra-violet filter (excitation wavelength = 330-380 nm). Labeled RGCs were counted along the median line of each quadrants starting from the optic disc to the peripheral border of the retina at 500 mm intervals, under an eyepiece grid of 200 X 200 mm. The percentage of surviving RGCs resulting from each treatment was expressed by comparing the number of surviving RGCs in the injured eyes with their contra-lateral eyes. All data were expressed as mean ± SEM. Statistical significance was evaluated by one way ANOVA, followed by a Tukey-Kramer post hoc test. Differences were considered significant for $p<0.05$. Sp35-Fc treated animals showed significant neuronal survival (83%) when compared to control-Fc or PBS treated animals, which each only showed approximately 50% neuronal survival.

Other Embodiments

[0192] Other embodiments are within the following claims.

Claims

1. An isolated nucleic acid comprising a nucleotide sequence encoding a polypeptide wherein: (a) the polypeptide comprises (i) an Sp35 LRR domain, (ii) an Sp35 basic region C-terminal to the LRR domain, and (iii) an Sp 35 immunoglobulin (Ig) domain C-terminal to the basic region; and (b) the polypeptide lacks a transmembrane domain.

2. An isolated nucleic acid comprising a nucleotide sequence encoding a polypeptide wherein the polypeptide comprises an Sp35 Ig domain and lacks an Sp35 LRR domain, an Sp35 basic region, a transmembrane domain, and a cytoplasmic domain.

3. An isolated nucleic acid comprising a nucleotide sequence encoding a polypeptide wherein the polypeptide comprises an Sp35 LRR domain and lacks an Sp35 Ig domain, an Sp35 basic region, a transmembrane domain, and a cytoplasmic domain.

4. The nucleic acid of claim 1, wherein the Sp35 polypeptide lacks a cytoplasmic domain.

5. The nucleic acid of claim 1, wherein the polypeptide comprises amino acid residues 34-532 of SEQ ID NO: 2.

6. The nucleic acid of any one of claims 1-3, wherein the polypeptide is a fusion polypeptide comprising a non-Sp35 moiety.

7. The nucleic acid of claim 6, wherein the non-Sp35 moiety is selected from the group consisting of an Ig moiety, a serum albumin moiety, a targeting moiety, a reporter moiety, and a purification-facilitating moiety.

8. The nucleic acid of claim 7, wherein the non-Sp35 moiety is an Ig moiety.

9. The nucleic acid of claim 8, wherein the Ig moiety is an Fc moiety.

10. The nucleic acid of any one of claims 1-3, wherein the nucleotide sequence is operatively linked to an expression control sequence.

11. A vector comprising the nucleic acid of claim 10.

12. A host cell comprising the vector of claim 11.

13. An isolated polypeptide, wherein: (a) the polypeptide comprises (i) an Sp35 LRR domain, (ii) an Sp35 basic region C-terminal to the LRR domain, and (iii) an immunoglobulin (Ig) domain C-terminal to the basic region; and (b) the polypeptide lacks a transmembrane domain.

14. An isolated polypeptide wherein the polypeptide comprises an Sp35 Ig domain and lacks an LRR domain, an Sp35 basic region, a transmembrane domain, and a cytoplasmic domain.

15. An isolated polypeptide wherein the polypeptide comprises an Sp35 LRR domain and lacks an Sp35 Ig domain, an Sp35 basic region, a transmembrane domain, and a cytoplasmic domain.

16. The polypeptide of claim 13, wherein one of the leucine rich repeats is a carboxy-terminal leucine-rich repeat (LRRCT).

17. The polypeptide of claim 13, wherein one of the leucine rich repeats is an amino-terminal leucine-rich repeat (LRRNT).

18. The polypeptide of claim 13, wherein the Sp35 polypeptide lacks a cytoplasmic domain.

19. The polypeptide of claim 13, wherein the polypeptide comprises amino acid residues 34-532 of SEQ ID NO: 2.

20. The polypeptide of claim 13, 14 or 15, wherein the polypeptide is a fusion polypeptide comprising a non-Sp35 moiety.

21. The polypeptide of claim 20, wherein the non-Sp35 moiety is selected from the group consisting of an Ig moiety, a serum albumin moiety, a targeting moiety, a reporter moiety, and a purification-facilitating moiety.

22. The polypeptide of claim 21, wherein the non-Sp35 moiety is an Ig moiety.

23. The polypeptide of claim 22, wherein the Ig moiety is an Fc moiety.

24. The polypeptide of claim 13, 14 or 15, wherein the polypeptide is conjugated to a polymer.

25. The polypeptide of claim 24, wherein the polymer is selected from the group consisting of a polyalkylene glycol, a sugar polymer, and a polypeptide.

26. The polypeptide of claim 25, wherein the polymer is a polyalkylene glycol.

27. The polypeptide of claim 26, wherein the polyalkylene glycol is polyethylene glycol (PEG).

28. The polypeptide of claim 24, wherein the polypeptide is conjugated to 1, 2, 3 or 4 polymers.

29. The polypeptide of claim 28, wherein the total molecular weight of the polymers is from 20,000 Da to 40,000 Da.

30. A method of inhibiting signal transduction by NgR1, comprising contacting the NgR1 with an effective amount of an Sp35 polypeptide.

31. A method of decreasing inhibition of axonal growth of a central nervous system (CNS) neuron, comprising contacting the neuron with an effective amount of a polypeptide selected from the group consisting of an Sp35 polypeptide, an anti-Sp35 antibody, or an antigen-binding fragment of an anti-Sp35 antibody.

32. A method of inhibiting growth cone collapse of a CNS neuron, comprising contacting the neuron with an effective amount of a polypeptide selected from the group consisting of an Sp35 polypeptide, an anti-Sp35 antibody, or an antigen-binding fragment of an anti-Sp35 antibody.

33. A method of treating a CNS disease, disorder or injury in a mammal, comprising administering to the mammal a therapeutically effective amount of a polypeptide selected from the group consisting of an Sp35 polypeptide, an anti-Sp35 antibody, or an antigen-binding fragment of an anti-Sp35 antibody.

34. The method of any one of claims 30-33, wherein the Sp35 polypeptide is selected from the group consisting of

a polypeptide, wherein: (a) the polypeptide comprises (i) an LRR domain comprising 12-14 Sp35 leucine-rich repeats, (ii) an Sp35 basic region C-terminal to the LRR domain, and (iii) an immunoglobulin (Ig) domain C-terminal to the basic region; and (b) the polypeptide lacks a transmembrane domain; and

a polypeptide wherein the polypeptide comprises an Sp35 Ig domain and lacks an LRR domain, a basic region, a transmembrane domain, and a cytoplasmic domain.

35. The method of claim 33, wherein the CNS disease, disorder or injury is a spinal cord injury or an optic nerve injury.

36. The method of claim 33, wherein the polypeptide is administered locally.

37. The method of claim 33, wherein the polypeptide is administered initially within 48 hours of the spinal cord injury.

38. The method of claim 36, wherein the therapeutically effective amount of the polypeptide is from 10 μg to 10 mg.

39. A method of treating a CNS disease, disorder or injury in a mammal, comprising (a) providing a cultured host cell expressing a recombinant Sp35 polypeptide; and (b) introducing the host cell into the mammal at or near the site of the CNS disease, disorder or injury.

40. The method of claim 39, wherein the disease, disorder or injury is a spinal cord injury.

41. The method of claim 39, wherein the cultured host cell is derived from the mammal to be treated.

42. The method of claim 39, wherein the recombinant Sp35 polypeptide is a full-length Sp35 polypeptide.

43. A method of promoting myelination at the site of the CNS disease, disorder or injury, comprising contacting the site of the CNS disease, disorder or injury with an effective amount of an Sp35 polypeptide.

44. The method of claim 43, wherein the Sp35 polypeptide comprises an Sp35 LRR domain and lacks an Sp35 Ig domain, an Sp35 basic region, a transmembrane domain, and a cytoplasmic domain.

45. A method of treating a CNS disease, disorder or injury by *in vivo* gene therapy, comprising administering to a mammal, at or near the site of the disease, disorder or injury, a viral vector comprising a nucleotide sequence that encodes an Sp35 polypeptide so that the Sp35 polypeptide is expressed from the nucleotide sequence in the mammal in an amount sufficient to reduce inhibition of axonal extension by neurons at or near the site of the injury.

46. The method of claim 36, wherein the viral vector is selected from the group consisting of an adenoviral vector, a lentiviral vector, a baculoviral vector, an Epstein Barr viral vector, a papovaviral vector, a vaccinia viral vector, and a herpes simplex viral vector.

47. The method of claim 45, wherein the disease, disorder or injury is selected from the group consisting of spinal cord injury and optic nerve injury.

48. The method of claim 45, wherein the viral vector is administered by a route selected from the group consisting of topical administration, intraocular administration, parenteral administration, intrathecal administration, subdural administration and subcutaneous administration.

49. A nucleic acid encoding a polypeptide comprising an Sp35 LRR domain, basic region, Ig domain, connecting sequence, and transmembrane domain; and lacking a functional cytoplasmic domain.

50. The nucleic acid of claim 49, wherein the nucleic acid encodes a polypeptide consisting essentially of amino acids 1-576 of SEQ ID NO: 2.

51. A method of promoting survival of a neuron at risk of dying, comprising contacting the neuron with an effective amount of an Sp35 polypeptide.

52. The method of claim 51, wherein the neuron is *in vitro*.

53. The method of claim 51, wherein the neuron is in a mammal with a neurodegenerative disease disorder or injury.

54. The method of claim 53, wherein the neurodegenerative disease, disorder or injury is selected from the group consisting of multiple sclerosis, ALS, Huntington's disease, Alzheimer's disease, Parkinson's disease, diabetic neuropathy, stroke, traumatic brain injuries and spinal cord injury.

55. A method of promoting survival of a neuron at risk of death of dying in a mammal with a neurodegenerative disease, disorder or injury, comprising (a) providing a cultured host cell expressing a recombinant Sp35 polypeptide; and (b) introducing the host cell into the mammal at the site of the neuron.

56. An *in vivo* gene therapy method of promoting survival of a neuron at risk of dying, comprising administering to a mammal, at or near the site of the neuron, a viral vector comprising a nucleotide sequence that encodes an Sp35 polypeptide so that the Sp35 polypeptide is expressed from the nucleotide sequence in the mammal in an amount sufficient to promote survival of the neuron.

57. The method of claim 51, wherein the Sp35 polypeptide is soluble.

FIG. 1

GGAGAGACATGCGATTGGTGACCGAGCCGAGCGGACCGAAGGCGCGCCCGA
GATGCAGGTGAGCAAGAGGATGCTGGCGGGGGGCGTGAGGAGCATGCCCAG
CCCCCTCCTGGCCTGCTGGCAGCCCATCCTCCTGCTGGTGCTGGGCTCAGTGC
TGTCAGGCTCGGCCACGGGCTGCCCGCCCCGCTGCGAGTGCTCCGCCCAGGA
CCGCGCTGTGCTGTGCCACCGCAAGCGCTTTGTGGCAGTCCCCGAGGGCATC
CCCACCGAGACGCGCCTGCTGGACCTAGGCAAGAACCGCATCAAAACGCTCA
ACCAGGACGAGTTCGCCAGCTTCCCGCACCTGGAGGAGCTGGAGCTCAACGA
GAACATCGTGAGCGCCGTGGAGCCCGGCGCCTTCAACAACCTCTTCAACCTC
CGGACGCTGGGTCTCCGCAGCAACCGCCTGAAGCTCATCCCGCTAGGCGTCT
TCACTGGCCTCAGCAACCTGACCAAGCTGGACATCAGCGAGAACAAGATTGT
TATCCTACTGGACTACATGTTTCAGGACCTGTACAACCTCAAGTCACTGGAGG
TTGGCGACAATGACCTCGTCTACATCTCTCACCGCGCCTTCAGCGGCCTCAAC
AGCCTGGAGCAGCTGACGCTGGAGAAATGCAACCTGACCTCCATCCCCACCG
AGGCGCTGTCCCACCTGCACGGCCTCATCGTCCTGAGGCTCCGGCACCTCAA
CATCAATGCCATCCGGGACTACTCCTTCAAGAGGCTCTACCGACTCAAGGTCT
TGGAGATCTCCCACTGGCCCTACTTGGACACCATGACACCCAACTGCCTCTAC
GGCCTCAACCTGACGTCCCTGTCCATCACACACTGCAATCTGACCGCTGTGCC
CTACCTGGCCGTCCGCCACCTAGTCTATCTCCGCTTCCTCAACCTCTCCTACA
ACCCCATCAGCACCATTGAGGGCTCCATGTTGCATGAGCTGCTCCGGCTGCA
GGAGATCCAGCTGGTGGGCGGGCAGCTGGCCGTGGTGGAGCCCTATGCCTTC
CGCGGCCTCAACTACCTGCGCGTGCTCAATGTCTCTGGCAACCAGCTGACCA
CACTGGAGGAATCAGTCTTCCACTCGGTGGGCAACCTGGAGACACTCATCCT
GGACTCCAACCCGCTGGCCTGCGACTGTCGGCTCCTGTGGGTGTTCCGGCGCC
GCTGGCGGCTCAACTTCAACCGGCAGCAGCCCACGTGCGCCACGCCCGAGTT
TGTCCAGGGCAAGGAGTTCAAGGACTTCCCTGATGTGCTACTGCCCAACTACT
TCACCTGCCGCCGCGCCCGCATCCGGGACCGCAAGGCCCAGCAGGTGTTTGT
GGACGAGGGCCACACGGTGCAGTTTGTGTGCCGGGCCGATGGCGACCCGCCG
CCCGCCATCCTCTGGCTCTCACCCCGAAAGCACCTGGTCTCAGCCAAGAGCA
ATGGGCGGCTCACAGTCTTCCCTGATGGCACGCTGGAGGTGCGCTACGCCCA
GGTACAGGACAACGGCACGTACCTGTGCATCGCGGCCAACGCGGGCGGCAA
CGACTCCATGCCCGCCCACCTGCATGTGCGCAGCTACTCGCCCGACTGGCCCC
ATCAGCCCAACAAGACCTTCGCTTTCATCTCCAACCAGCCGGGCGAGGGAGA
GGCCAACAGCACCCGCGCCACTGTGCCTTTCCCCTTCGACATCAAGACCCTCA
TCATCGCCACCACCATGGGCTTCATCTCTTTCCTGGGCGTCGTCCTCTTCTGCC
TGGTGCTGCTGTTTCTCTGGAGCCGGGGCAAGGGCAACACAAAGCACAACAT
CGAGATCGAGTATGTGCCCCGAAAGTCGGACGCAGGCATCAGCTCCGCCGAC
GCGCCCCGCAAGTTCAACATGAAGATGATATGAGGCCGGGGCGGGGGGCAG
GGACCCCCGGGCGGCCGGGCAGGGGAAGGGGCCTGGCCGCCACCTGCTCACT
CTCCAGTCCTTCCCACCTCCTCCCTACCCTTCTACACACGTTCTCTTTCTCCCT
CCCGCCTCCGTCCCCTGCTGCCCCCCGCCAGCCCTCACCACCTGCCCTCCTTC
TACCAGGACCTCAGAAGCCCAGACCTGGGGACCCCACCTACACAGGGGCATT
GACAGACTGGAGTTGAAAGCCGACGAACCGACACGCGGCAGAGTCAATAAT
TCAATAAAAAAGTTACGAACTTTCTCTGTAACTTGGGTTTCAATAATTATGGA
TTTTTATGAAAACTTGAAATAATAAAAAGAGAAAAAAACTATTTCCTATAGC

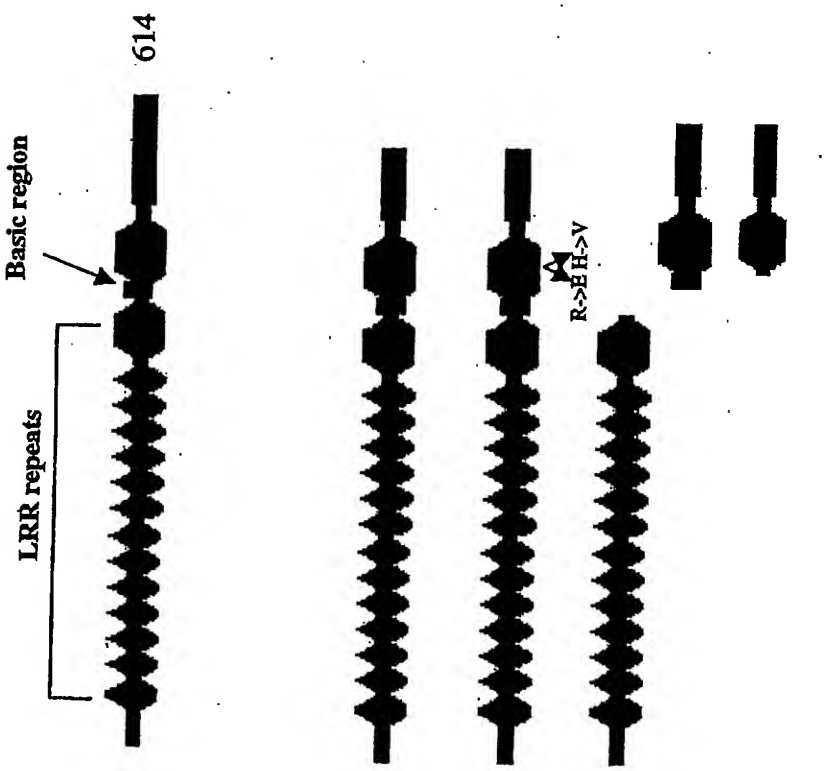
FIG. 1 (continued)

TAGTCGGAATGCAAACTTTTGACGTCCTGATTGCTCCAGGGCCCTCTTCCAAC
TCAGTTTCTTGTTTTTCTCTTCNTCCTNCTCCTCTTCTTCCTCCTTTCTCTTCTCT
TCCCCCAGTGGGGAGGGATCACTCAGGAAAACAGGAAAGGAGGTTCCAGCC
CCACCCACCTGCCCACCCCGCCCCAGGCACCATCAGGAGCAGGCTAGGGGGC
AGGCCTGGGCCCAGCTCCGGGCTGGCTTTTTGCAGGGCGCAGGTGGAGGGGA
CAGGTCTGCCGATGGGGGTGGGAGCCTGTCTGCTGGGCTGCCAGGCGGCACC
ACTGCAAGGGGTGGGAGCCTGGCTCGGGTGTGGCTGAGACTCTGGACAGAGG
CTGGGGTCCTCCTGGGGGACAGCACAGTCAGTGGAGAGAGCCAGGGGCTGG
AGGTGGGGCCCACCCCAGCCTCTGGTCCCAGCTCTGCTGCTCACTTGCTGTGT
GGCCCTCAAGCAGGTCCACTGGCCTCTCTGGGCCTCAGTCTCCACATCTGTAC
AAATGGGAACATTACCCCCTGCCCTGCCTACCTNANAGGGCTGTTNTGAGGN
ATNGATGAGATGATGTATGT

FIG. 2

MLAGGVRSMPSPLLACWQPILLLVLGSVL
SGSATGCPPRCECSAQDRAVLCHRKRFVA
VPEGIPTETRLLDLGKNRIKTLNQDEFASF
PHLEELELNENIVSAVEPGAFNNLFNLRTL
GLRSNRLKLIPLGVFTGLSNLTKLDISENKI
VILLDYMFQDLYNLKSLEVGDNDLVYISHR
AFSGLNSLEQLTLEKCNLTSIPTEALSHLH
GLIVLRLRHLNINAIRDYSFKRLYRLKVLEI
SHWPYLDTMTPNCLYGLNLTSLSITHCNLT
AVPYLAVRHLVYLRFLNLSYNPISTIEGSM
LHELLRLQEIQLVGGQLAVVEPYAFRGLNY
LRVLNVSGNQLTTLEESVFHSVGNLETLIL
DSNPLACDCRLLWVFRRRWRLNFNRQQPT
CATPEFVQGKEFKDFPDVLLPNYFTCRRA
RIRDRKAQQVFVDEGHTVQFVCRADGDPP
PAILWLSPRKHLVSAKSNGRLTVFPDGTLE
VRYAQVQDNGTYLCIAANAGGNDSMPAHL
HVRSYSPDWPHQPNKTFAFISNQPGEGEA
NSTRATVPFPFDIKTLIIATTMGFISFLGVV
LFCLVLLFLWSRGKGNTKHNIEIEYVPRKS
DAGISSADAPRKFNMKMI

FIG. 3



Deletion Construct (amino acids)	EC 50
(34-532)	6 nm
(34-532) mutated	120 nm
(34-416)	120 nm
(417-532)	60 nm
(425-532)	25 nm

FIG. 4

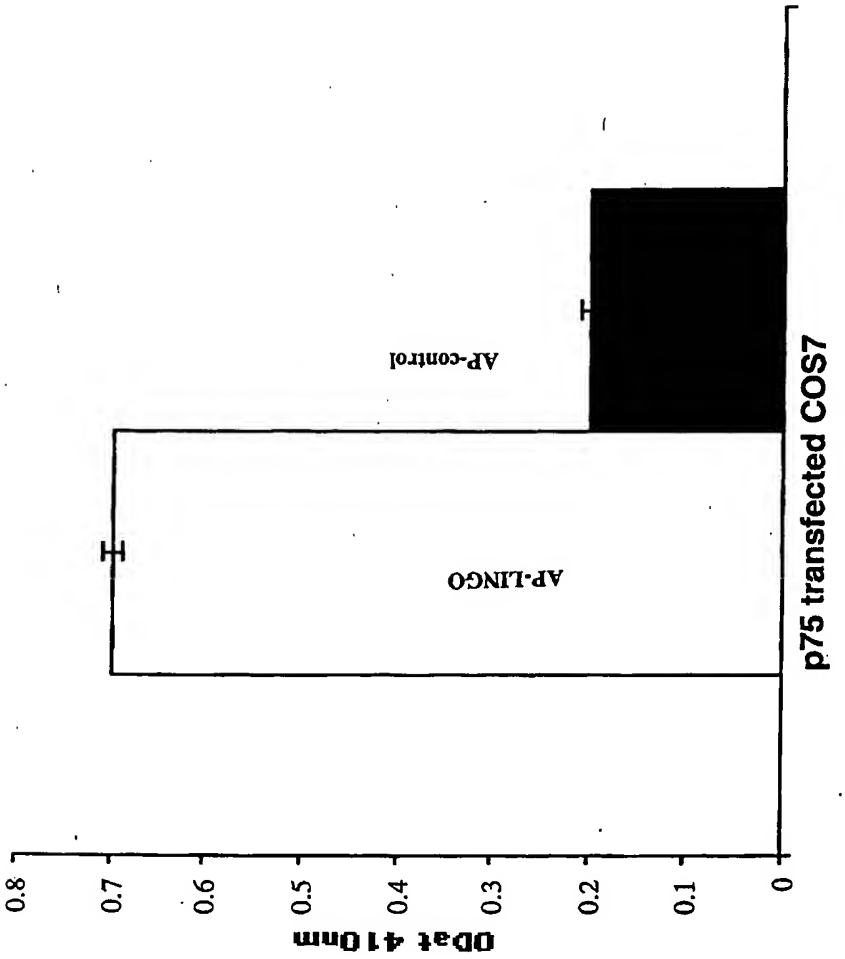

p75 transfected COS7
AP-LINGO
AP-control
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
ODat 410nm

FIG. 5

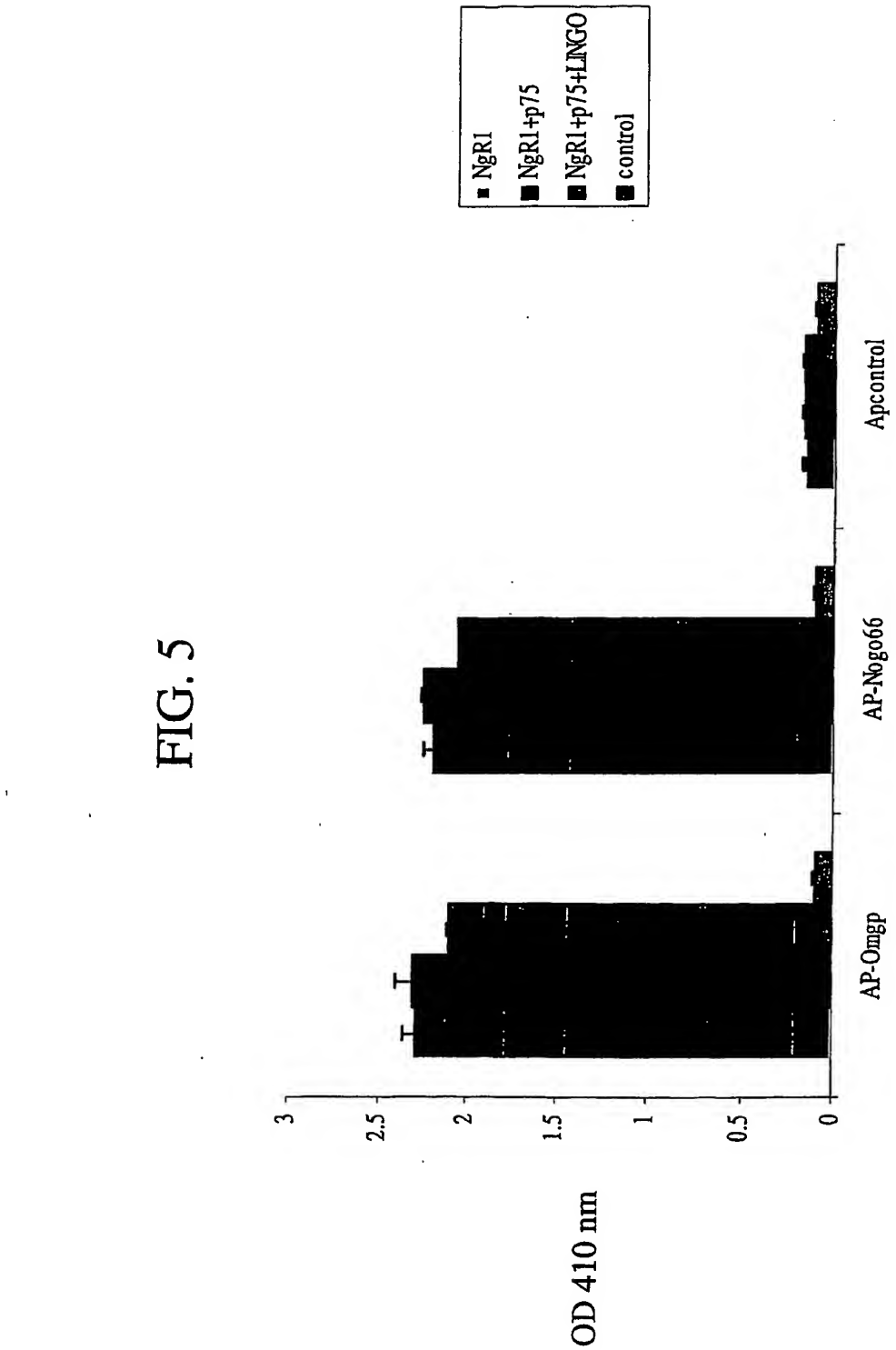

NgR1
NgR1+p75
NgR1+p75+LINGO
control
OD 410 nm
3
2.5
2
1.5
1
0.5
0
AP-Omgp
AP-Nogo66
Apcontrol

FIG. 6

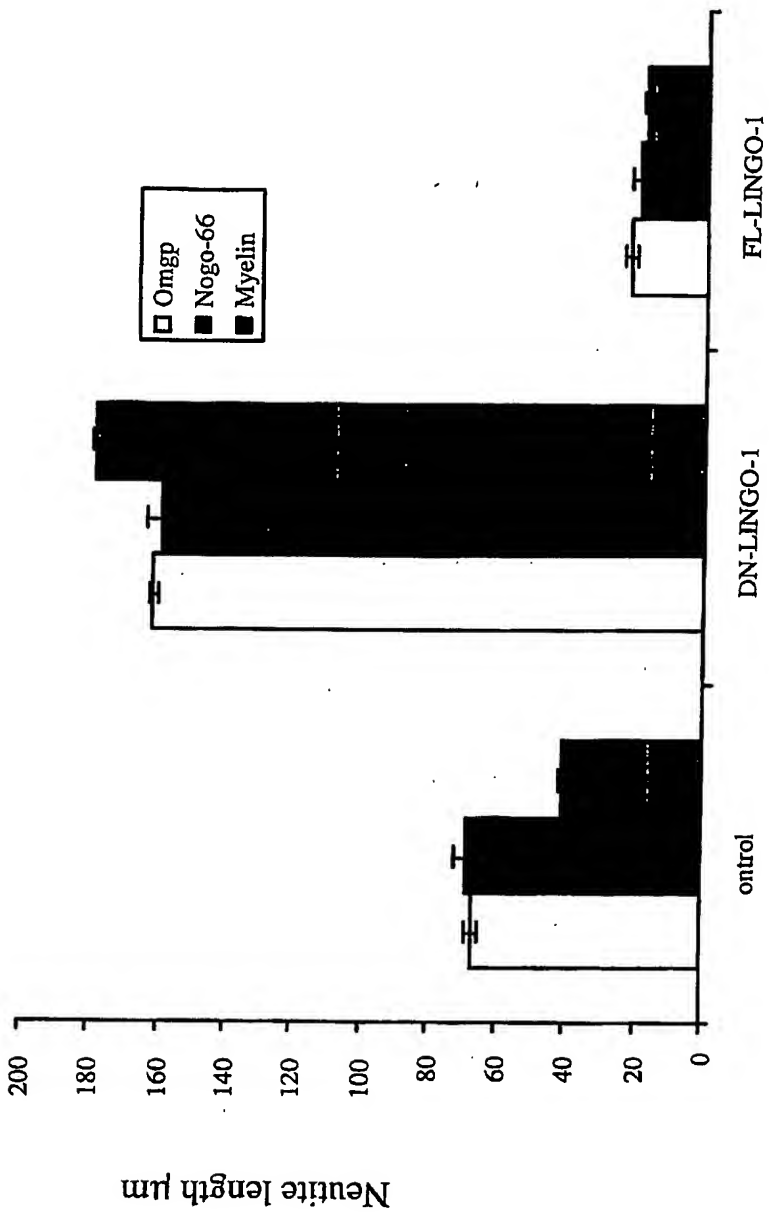

Omgp
Nogo-66
Myelin
Neutite length µm
200
180
160
140
120
100
80
60
40
20
0
ontrol
DN-LINGO-1
FL-LINGO-1

FIG. 7

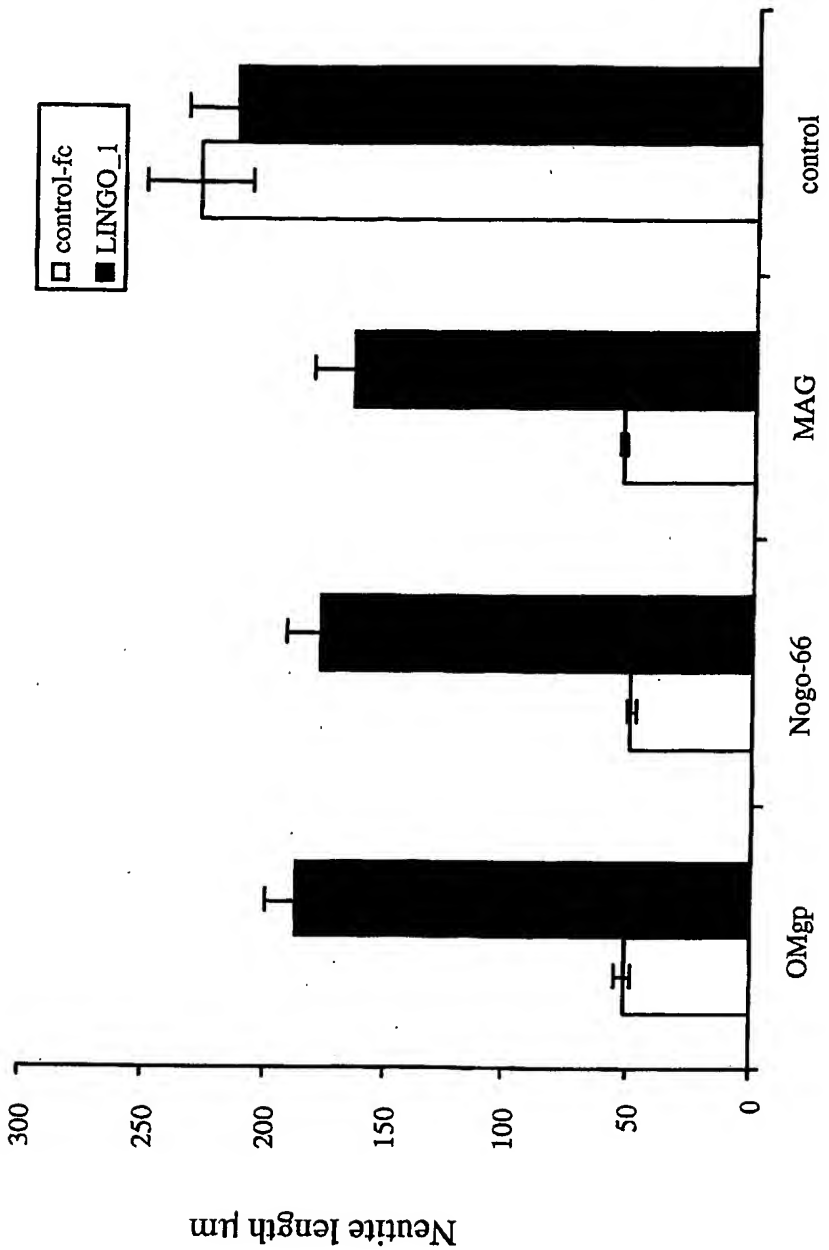

control-fc
LINGO_1
Neurite length µm
300
250
200
150
100
50
0
OMgp
Nogo-66
MAG
control

FIG. 9

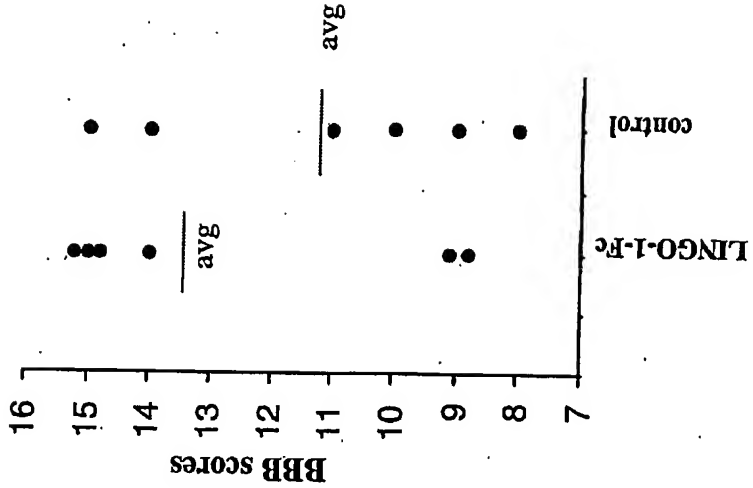
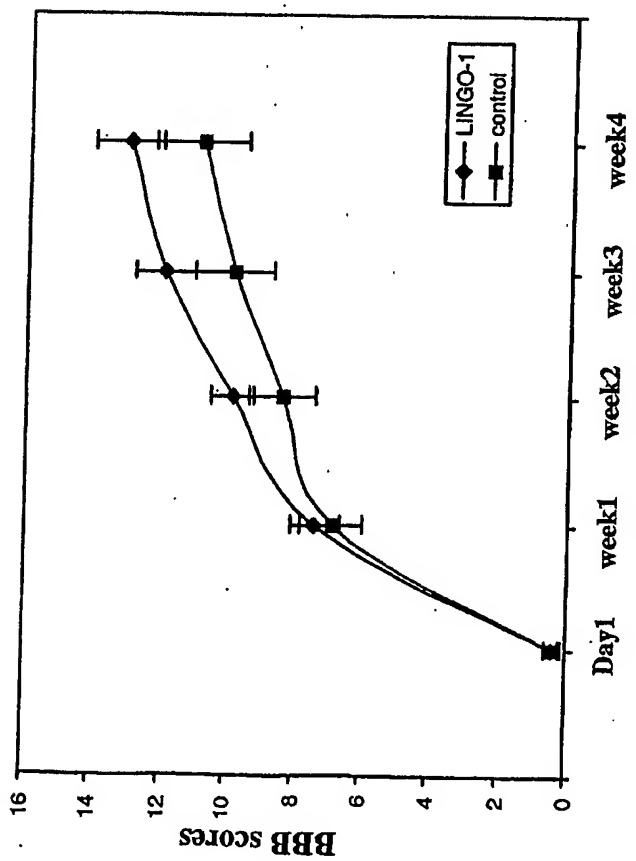

BBB scores
16
15
14
13
12
11
10
9
8
7
avg
avg
LINGO-1-Fc
control

FIG. 8


BBB scores
16
14
12
10
8
6
4
2
0
LINGO-1
control
Day1
week1
week2
week3
week4

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
7 October 2004 (07.10.2004)

PCT

(10) International Publication Number
WO 2004/085648 A3

(51) International Patent Classification[7]: C12N 15/12, 15/62, 15/63, C07K 14/47, 16/18, A61K 38/17, 39/395, 48/00

(21) International Application Number: PCT/US2004/008323

(22) International Filing Date: 17 March 2004 (17.03.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/455,756	19 March 2003 (19.03.2003)	US
60/480,241	20 June 2003 (20.06.2003)	US
60/492,057	1 August 2003 (01.08.2003)	US

(71) Applicant *(for all designated States except US)*: **BIOGEN IDEC MA INC.** [US/US]; 14 Cambridge Center, Cambridge, MA 02142 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **MI, SHA** [US/US]; 4 Vernon Road, Belmont, MA 02478 (US). **MCCOY, JOHN** [GB/US]; 56 Howard Street, Reading, MA 01867 (US). **PEPINSKY, R., BLAKE** [US/US]; 30 Falmouth Road, Arlington, MA 02474 (US). **LEE, DANIEL, H., S.** [CA/US]; 361 Dutton Road, Sudbury, MA 01776 (US).

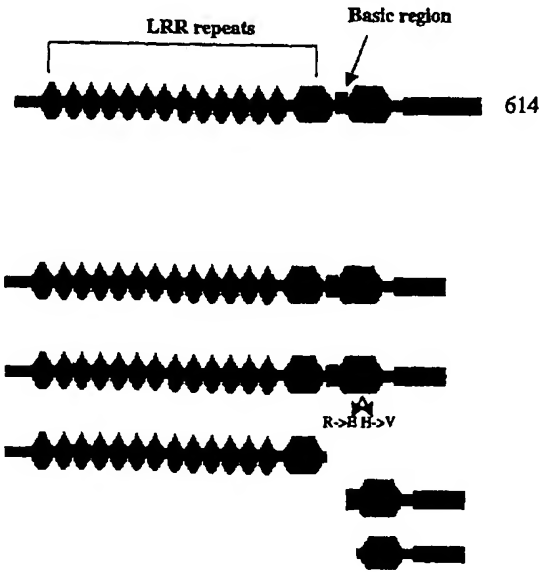
(74) Agents: **HALEY, JAMES, F., JR.** et al.; c/o Fish & Neave, 1251 Avenue of the Americas, New York, NY 10020 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK,

[Continued on next page]

(54) Title: NOGO RECEPTOR BINDING PROTEIN



Deletion Construct (amino acids)	EC 50
(34-532)	6 nm
(34-532) mutated	120 nm
(34-416)	120 nm
(417-532)	60 nm
(425-532)	25 nm

(57) Abstract: The invention provides Sp35 polypeptides and fusion proteins thereof, Sp35 antibodies and antigen-binding fragments thereof and nucleic acids encoding the same. The invention also provides compositions comprising, and methods for making and using, such Sp35 antibodies, antigen-binding fragments thereof, Sp35 polypeptides and fusion proteins thereof.

WO 2004/085648 A3

TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

(88) Date of publication of the international search report:
18 November 2004

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

INTERNATIONAL SEARCH REPORT

International Application No
T/US2004/008323

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/12 C12N15/62 C12N15/63 C07K14/47 C07K16/18
A61K38/17 A61K39/395 A61K48/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, BIOSIS, EMBASE, EMBL, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	LI W ET AL: "NEUTRALIZATION OF MYELIN - ASSOCIATED NOGO - A BY A NOGO RECEPTOR - Fc FUSION PROTEIN" SOCIETY FOR NEUROSCIENCE ABSTRACTS, SOCIETY FOR NEUROSCIENCE, US, 2002, page ABS3332, XP001199824 ISSN: 0190-5295 the whole document	1-57
A	WO 01/12662 A (INCYTE GENOMICS INC ; PATTERSON CHANDRA (US); AZIMZAI YALDA (US); YUE) 22 February 2001 (2001-02-22) MEPAP-11 (SEQ ID NO: 11)	1-57

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
29 September 2004	15/10/2004

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Barnas, C

Form PCT/ISA/210 (second sheet) (January 2004)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	MI SHA ET AL: "LINGO-1 is a component of the Nogo-66 receptor/p75 signaling complex." March 2004 (2004-03), NATURE NEUROSCIENCE, VOL. 7, NR. 3, PAGE(S) 221-228 , XP001183275 ISSN: 1097-6256 figures 1,5,6	1-57
P,A	MI S ET AL: "A novel CNS - specific protein promotes axonal elongation by modulating RhoA signaling." 2003, SOCIETY FOR NEUROSCIENCE ABSTRACT VIEWER AND ITINERARY PLANNER, VOL. 2003, PAGE(S) ABSTRACT NO. 891.5 URL - HTTP://SFN.SCHOLARONE.COM , 33RD ANNUAL MEETING OF THE SOCIETY OF NEUROSCIENCE; NEW ORLEANS, LA, USA; NOVEMBER 08-12, 2003 , XP001183276 sentences 1-13	1-57

INTERNATIONAL SEARCH REPORT

nternational application No.

PCT/US2004/008323

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.b of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application and necessary to the claimed invention, the international search was carried out on the basis of:

   a. type of material

   [X] a sequence listing

   [ ] table(s) related to the sequence listing

   b. format of material

   [X] in written format

   [X] in computer readable form

   c. time of filing/furnishing

   [ ] contained in the international application as filed

   [ ] filed together with the international application in computer readable form

   [X] furnished subsequently to this Authority for the purpose of search

2. [X] In addition, in the case that more than one version or copy of a sequence listing and/or table relating thereto has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that in the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

INTERNATIONAL SEARCH REPORT

nternational application No.
PCT/US2004/008323

Box II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.: 30-48, 51-57 (part)
because they relate to subject matter not required to be searched by this Authority, namely:

Rule 39.1(iv) PCT - Method for treatment of the human or animal body by therapy

2. [ ] Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. [ ] As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. [ ] As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. [ ] No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
T/US2004/008323

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 0112662 A	22-02-2001	AU 6906800 A	13-03-2001
		CA 2382015 A1	22-02-2001
		EP 1206543 A2	22-05-2002
		JP 2003527089 T	16-09-2003
		WO 0112662 A2	22-02-2001
		US 2002182671 A1	05-12-2002
		US 2003124649 A1	03-07-2003

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C07H 21/02, 21/04, 1/00, 17/00, C12Q 1/68, G01N 33/53	A1	(11) International Publication Number: **WO 99/06427** (43) International Publication Date: 11 February 1999 (11.02.99)

(21) International Application Number: PCT/US98/16241

(22) International Filing Date: 4 August 1998 (04.08.98)

(30) Priority Data:
60/054,645 4 August 1997 (04.08.97) US

(71) Applicant: MILLENNIUM BIOTHERAPEUTICS, INC. [US/US]; 620 Memorial Drive, Cambridge, MA 02142 (US).

(72) Inventor: MCCARTHY, Sean, A.; 62 Commonwealth Avenue #4, Boston, MA 02116 (US).

(74) Agent: MEIKLEJOHN, Anita, L.; Fish & Richardson P.C., 225 Franklin Street, Boston, MA 02110–2804 (US).

(81) Designated States: AU, CA, JP, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: TANGO–78, TANGO–79, AND TANGO–81 NUCLEIC ACID MOLECULES AND POLYPEPTIDES

(57) Abstract

The invention relates to Tango–78, Tango–79, and Tango–81 polypeptides, nucleic acid molecules encoding Tango–78, Tango–79, and Tango–81, and uses thereof.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

TANGO-78, TANGO-79, AND TANGO-81 NUCLEIC ACID MOLECULES AND POLYPEPTIDES

Summary of the Invention

The invention relates to the discovery and characterization of the genes encoding Tango-78, Tango-79, and Tango-81.

The invention features isolated nucleic acid molecules encoding Tango-78, Tango-79, or Tango-81, the isolated nucleic acid molecules that encode polypeptides that are substantially identical to the Tango-78, Tango-79, or Tango-81 protein sequences described herein (SEQ ID NOS:2, 4, or 6) and isolated nucleic acid molecules which hybridize under stringent conditions to the protein coding portions of the Tango-78, Tango-79, or Tango-81 nucleic acid molecules described herein.

The invention also features a host cell which includes an isolated nucleic acid molecule encoding Tango-78, Tango-79, or Tango-81, a nucleic acid vector (e.g., an expression vector; a vector which includes a regulatory element; a vector which includes a regulatory element selected from the group consisting of the cytomegalovirus hCMV immediate early gene, the early promoter of SV40 adenovirus, the late promoter of SV40 adenovirus, the lac system, the trp system, the TAC system, the TRC system, the major operator and promoter regions of phage λ, the control regions of fd coat protein, the promoter for 3-phosphoglycerate kinase, the promoters of acid phosphatase, and the promoters of the yeast α-mating factors; a vector which includes a regulatory element which directs tissue-specific expression; a vector which includes a reporter gene; a vector which includes a reporter gene selected from the group selected from the group consisting of β-lactamase, chloramphenicol acetyltransferase (CAT), adenosine deaminase (ADA), aminoglycoside phosphotransferase (neo^r,

G418^r), dihydrofolate reductase (DHFR), hygromycin-B-phosphotransferase (HPH), thymidine kinase (TK), lacZ (encoding β-galactosidase), and xanthine guanine phosphoribosyltransferase (XGPRT); a vector that is a plasmid; a vector that is a virus; a vector that is a retrovirus.

In another embodiment, the invention features a substantially pure Tango-78, Tango-79, or Tango-81 polypeptide (e.g., a Tango-78, Tango-79, or Tango-81 polypeptide that is soluble under physiological conditions; a Tango-78, Tango-79, or Tango-81 polypeptide which includes a signal sequence; a Tango-78 polypeptide that is at least 85%, 90%, 95%, or 100% identical to the amino acid sequence of SEQ ID NO:2; a Tango-79 polypeptide that is at least 85%, 90%, 95%, or 100% identical to the amino acid sequence of SEQ ID NO:4; and a Tango-81 polypeptide that is at least 85%, 90%, 95%, or 100% identical to the amino acid sequence of SEQ ID NO:6.

In other embodiments, the invention features a substantially pure polypeptide which includes a first portion and a second portion, the first portion including a Tango-78, Tango-79, or Tango-81 polypeptide and the second portion including a detectable marker.

The invention also features antibodies, e.g., monoclonal antibodies, that selectively binds to a polypeptide of the invention (Tango-78, Tango-79, or Tango-81).

The invention also features a pharmaceutical composition which includes a Tango-78, Tango-79, or Tango-81 polypeptide.

The invention also features a method for diagnosing a disorder associated with aberrant expression of Tango-78 the method including obtaining a biological sample from a patient and measuring Tango-78 expression in the biological sample, wherein increased or decreased

Tango-78 expression in the biological sample compared to a control indicates that the patient suffers from a disorder associated with aberrant expression of Tango-78.

The invention also features a method for diagnosing a disorder associated with aberrant expression of Tango-79, the method including obtaining a biological sample from a patient and measuring Tango-79 expression in the biological sample, wherein increased or decreased Tango-79 expression in the biological sample compared to a control indicates that the patient suffers from a disorder associated with aberrant expression of Tango-79.

The invention also features a method for diagnosing a disorder associated with aberrant expression of Tango-81, the method including obtaining a biological sample from a patient and measuring Tango-81 expression in the biological sample, wherein increased or decreased Tango-81 expression in the biological sample compared to a control indicates that the patient suffers from a disorder associated with aberrant expression of Tango-81.

The invention encompasses isolated nucleic acid molecules encoding Tango-78, Tango-79, or Tango-81 or a polypeptide fragment thereof; vectors containing these nucleic acid molecules; cells harboring recombinant DNA encoding Tango-78, Tango-79, or Tango-81; fusion proteins which include Tango-78, Tango-79, or Tango-81; transgenic animals which express Tango-78, Tango-79, or Tango-81; recombinant knock-out animals which fail to express Tango-78, Tango-79, or Tango-81.

The invention encompasses nucleic acids that have a sequence that is substantially identical to the nucleic acid sequence of Tango-78, Tango-79, or Tango-81. A nucleic acid sequence which is substantially identical to a given reference nucleic acid sequence is hereby defined as a nucleic acid having a sequence that has at least 85%, preferably 90%, and more preferably 95%, 98%, 99% or

more identity to the sequence of the given reference nucleic acid sequence, e.g., the nucleic acid sequence of SEQ ID NO:1, SEQ ID NO:3, or SEQ ID NO:5.

The invention encompasses polypeptides that have a sequence that is substantially identical to the amino acid sequence of Tango-78, Tango-79, or Tango-81. A polypeptide which is "substantially identical" to a given reference polypeptide is a polypeptide having a sequence that has at least 85%, preferably 90%, and more preferably 95%, 98%, 99% or more identity to the sequence of the given reference polypeptide sequence, e.g., the amino sequence of SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:6.

The nucleic acid molecules of the invention can be inserted into vectors, as described below, which will facilitate expression of the insert. The nucleic acid molecules and the polypeptides they encode can be used directly as diagnostic or therapeutic agents, or (in the case of a polypeptide) can be used to generate antibodies that, in turn, are therapeutically useful. Accordingly, expression vectors containing the nucleic acid molecules of the invention, cells transfected with these vectors, the polypeptides expressed, and antibodies generated (against either the entire polypeptide or an antigenic fragment thereof) are among the preferred embodiments.

A transformed cell is any cell into which (or into an ancestor of which) has been introduced, by means of recombinant DNA techniques, a nucleic acid encoding a polypeptide of the invention (e.g., a Tango-78, Tango-79, or Tango-81 polypeptide).

An isolated nucleic acid molecule is a nucleic acid molecule that is separated from the 5′ and 3′ coding sequences with which it is immediately contiguous in the naturally occurring genome of an organism. Isolated nucleic acid molecules include nucleic acid molecule

which are not naturally occurring, e.g., nucleic acid molecules created by recombinant DNA techniques.

Nucleic acid molecules include both RNA and DNA, including cDNA, genomic DNA, and synthetic (e.g., chemically synthesized) DNA. Where single-stranded, the nucleic acid molecule may be a sense strand or an antisense strand.

The invention also encompasses nucleic acid molecules that hybridize, preferably under stringent conditions, to a nucleic acid molecule encoding a Tango-78, Tango-79, or Tango-81 polypeptide (e.g., the polypeptide encoding portions of SEQ ID NO:1, SEQ ID NO:3, or SEQ ID NO:5). Preferably the hybridizing nucleic acid molecule consists of 400, more preferably 200 nucleotides. Preferred hybridizing nucleic acid molecules have a biological activity possessed by Tango-78, Tango-79, or Tango-81.

The invention also features substantially pure or isolated Tango-78, Tango-79, or Tango-81 polypeptides, including those that correspond to various functional domains of Tango-78, Tango-79, or Tango-81, or fragments thereof.

The polypeptides of the invention can be prepared by recombinant gene expression, chemically synthesized, or purified from tissues in which they are naturally expressed using standard biochemical methods of purification.

Also included in the invention are functional polypeptides, which possess one or more of the biological functions or activities of Tango-78, Tango-79, or Tango-81. These functions include the ability to bind some or all of the proteins which normally bind to Tango-78, Tango-79, or Tango-81. A functional polypeptide is also considered within the scope of the invention if it serves as an antigen for production of antibodies that

specifically bind to Tango-78, Tango-79, or Tango-81. In many cases, functional polypeptides retain one or more domains present in the naturally-occurring form of the polypeptide.

The functional polypeptides may contain a primary amino acid sequence that has been modified from those disclosed herein. Preferably these modifications consist of conservative amino acid substitutions, as described herein.

The terms "protein" and "polypeptide" are used herein interchangably to describe any chain of amino acids, regardless of length or post-translational modification (for example, glycosylation or phosphorylation). Thus, the term "Tango-78, Tango-79, or Tango-81 polypeptide" includes: full-length, naturally occurring Tango-78, Tango-79, or Tango-81 protein; recombinantly or synthetically produced polypeptide that corresponds to a full-length naturally occurring Tango-78, Tango-79, or Tango-81; or particular domains or portions of the naturally occurring protein. The term also encompasses mature Tango-78, Tango-79, or Tango-81 which has an added amino-terminal methionine (useful for expression in prokaryotic cells).

The term "purified" as used herein refers to a nucleic acid or peptide that is substantially free of cellular material, viral material, or culture medium when produced by recombinant DNA techniques, or chemical precursors or other chemicals when chemically synthesized.

Polypeptides or other compounds of interest are said to be "substantially pure" when they are within preparations that are at least 60% by weight (dry weight) the compound of interest. Preferably, the preparation is at least 75%, more preferably at least 90%, and most preferably at least 99%, by weight the compound of

interest. Purity can be measured by any appropriate standard method, for example, by column chromatography, polyacrylamide gel electrophoresis, or HPLC analysis.

Where a particular polypeptide or nucleic acid molecule is said to have a specific percent identity to a reference polypeptide or nucleic acid molecule of a defined length, the percent identity is relative to the reference polypeptide or nucleic acid molecule. Thus, a peptide that is 50% identical to a reference polypeptide that is 100 amino acids long can be a 50 amino acid polypeptide that is completely identical to a 50 amino acid long portion of the reference polypeptide. It might also be a 100 amino acid long polypeptide which is 50% identical to the reference polypeptide over its entire length. Of course, many other polypeptides will meet the same criteria. The same rule applies for nucleic acid molecules.

For polypeptides, the length of the reference polypeptide sequence will generally be at least 16 amino acids, preferably at least 20 amino acids, more preferably at least 25 amino acids, and most preferably 35 amino acids, 50 amino acids, or 100 amino acids. For nucleic acids, the length of the reference nucleic acid sequence will generally be at least 50 nucleotides, preferably at least 60 nucleotides, more preferably at least 75 nucleotides, and most preferably 100 nucleotides or 300 nucleotides.

In the case of polypeptide sequences which are less than 100% identical to a reference sequence, the non-identical positions are preferably, but not necessarily, conservative substitutions for the reference sequence. Conservative substitutions typically include substitutions within the following groups: glycine and alanine; valine, isoleucine, and leucine; aspartic acid and glutamic acid; asparagine and glutamine; serine and

threonine; lysine and arginine; and phenylalanine and tyrosine.

Sequence identity can be measured using sequence analysis software (for example, the Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, WI 53705), with the default parameters as specified therein.

The invention also features antibodies, e.g., monoclonal, polyclonal, and engineered antibodies, which specifically bind Tango-78, Tango-79, or Tango-81. By "specifically binds" is meant an antibody that recognizes and binds to a particular antigen, e.g., a Tango-78, Tango-79, or Tango-81 polypeptide of the invention, but which does not substantially recognize or bind to other molecules in a sample, e.g., a biological sample, which includes the polypeptide.

The invention also features antagonists and agonists of Tango-78, Tango-79, or Tango-81 that can inhibit or enhance, respectively, one or more of the biological activities of Tango-78, Tango-79, or Tango-81. Suitable antagonists can include small molecules (i.e., molecules with a molecular weight below about 500); large molecules (i.e., molecules with a molecular weight above about 500), antibodies that bind and "neutralize" Tango-78, Tango-79, or Tango-81 (as described below); polypeptides which compete with a native form of Tango-78, Tango-79, or Tango-81 for binding to a functional binding partner of the native protein; and nucleic acid molecules that interfere with transcription of Tango-78, Tango-79, or Tango-81 (for example, antisense nucleic acid molecules and ribozymes). Agonists of Tango-78, Tango-79, or Tango-81 also include small and large molecules, and antibodies other than neutralizing antibodies.

The invention also features molecules which can increase or decrease the expression of Tango-78, Tango-79, or Tango-81 (e.g., by influencing transcription or translation). Small molecules (i.e., molecules with a molecular weight below about 500), large molecules (i.e., molecules with a molecular weight above about 500), and nucleic acid molecules that can be used to inhibit the expression of Tango-78, Tango-79, or Tango-81 (for example, antisense and ribozyme molecules) or to enhance their expression (for example, molecules that bind to a Tango-78, Tango-79, or Tango-81 transcription regulatory sequence and increase transcription.

In addition, the invention features substantially pure polypeptides that functionally interact with Tango-78, Tango-79, or Tango-81 and the nucleic acid molecules that encode them.

The invention encompasses methods for treating disorders associated with aberrant expression or activity of a protein of the invention (i.e., Tango-78, Tango-79, or Tango-81). Thus, the invention includes methods for treating disorders associated with excessive expression or activity of the protein. Such methods entail administering a compound which decreases the expression of the protein. The invention also includes methods for treating disorders associated with insufficient expression or activity of a protein of the invention. These methods entail administering a compound which increases the expression or activity of the protein.

The invention also features methods for detecting a protein of the invention. Such methods include: obtaining a biological sample; contacting the sample with an antibody that specifically binds to the protein under conditions which permit specific binding; and detecting any antibody-protein complexes formed.

In addition, the present invention encompasses methods and compositions for the diagnostic evaluation, typing, and prognosis of disorders associated with inappropriate expression or activity of Tango-78, Tango-79, or Tango-81. For example, the nucleic acid molecules of the invention can be used as diagnostic hybridization probes to detect, for example, inappropriate expression of Tango-78, Tango-79, or Tango-81 or mutations in the Tango-78, Tango-79, or Tango-81 gene. Such methods may be used to classify cells by the level of Tango-78, Tango-79, or Tango-81 expression.

Thus, the invention features a method for diagnosing a disorder associated with aberrant activity of a protein of the invention, the method including obtaining a biological sample from a patient and measuring the activity of the protein in the biological sample, wherein increased or decreased activity in the biological sample compared to a control indicates that the patient suffers from a disorder associated with aberrant activity of the protein.

The nucleic acid molecules can be used as primers for diagnostic PCR analysis for the identification of gene mutations, allelic variations and regulatory defects in the Tango-78, Tango-79, or Tango-81 gene. The present invention further provides for diagnostic kits for the practice of such methods.

The invention features methods of identifying compounds that modulate the expression or activity of a protein of the invention by assessing the expression or activity of the protein in the presence and absence of a selected compound. A difference in the level of expression or activity of the protein in the presence and absence of the selected compound indicates that the selected compound is capable of modulating expression or activity of the protein. Expression can be assessed

either at the level of gene expression (e.g., by measuring mRNA) or protein expression by techniques that are well known to skilled artisans.

The preferred methods and materials are described below in examples which are meant to illustrate, not limit, the invention. Skilled artisans will recognize methods and materials that are similar or equivalent to those described herein, and that can be used in the practice or testing of the present invention.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods and materials are described herein. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In the case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and are not intended to be limiting.

Other features and advantages of the invention will be apparent from the detailed description, and from the claims.

Brief Description of the Drawing

Figure 1 is a depiction of the nucleic acid sequence (SEQ ID NO:1) and deduced amino acid sequence (SEQ ID NO:2) of Tango-78.

Figure 2 is a depiction of the nucleic sequence (SEQ ID NO:3) and deduced amino acid sequence (SEQ ID NO:4) of Tango-79.

Figure 3 is a depiction of the nucleic acid sequence (SEQ ID NO:5) and deduced amino acid sequence (SEQ ID NO:6) of Tango-81.

Figure 4 is an alignment of the amion acid sequence of Tango-78 and the amion acid sequence of murine nodal protein.

Figure 5 is an alignment between the amino acid sequence of Tango-79 and D45913 (Leucine rich repeat protein).

Figure 6 is a depiction of the results of Northern blot analysis of Tango-81 expression.

Detailed Description

Tango-78, Tango-79, and Tango-81 Nucleic Acid Molecules

The Tango-78, Tango-79, and Tango-81 nucleic acid molecules of the invention can be cDNA, genomic DNA, synthetic DNA, or RNA, and can be double-stranded or single-stranded (*i.e.*, either a sense or an antisense strand). Fragments of these molecules are also considered within the scope of the invention, and can be produced, for example, by the polymerase chain reaction (PCR) or generated by treatment with one or more restriction endonucleases. A ribonucleic acid (RNA) molecule can be produced by *in vitro* transcription.

The nucleic acid molecules of the invention can contain naturally occurring sequences, or sequences that differ from those that occur naturally, but, due to the degeneracy of the genetic code, encode the same polypeptide. In addition, these nucleic acid molecules are not limited to sequences that only encode polypeptides, and thus, can include some or all of the non-coding sequences that lie upstream or downstream from a coding sequence.

The nucleic acid molecules of the invention can be synthesized (for example, by phosphoramidite-based synthesis) or obtained from a biological cell, such as the cell of a mammal. Thus, the nucleic acids can be those of a human, mouse, rat, guinea pig, cow, sheep, horse, pig, rabbit, monkey, dog, or cat. Combinations or modifications of the nucleotides within these types of nucleic acids are also encompassed.

In addition, the isolated nucleic acid molecules of the invention encompass fragments that are not found as such in the natural state. Thus, the invention encompasses recombinant molecules, such as those in which a nucleic acid molecule (for example, an isolated nucleic acid molecule encoding Tango-78, Tango-79, or Tango-81) is incorporated into a vector (for example, a plasmid or viral vector) or into the genome of a heterologous cell (or the genome of a homologous cell, at a position other than the natural chromosomal location). Recombinant nucleic acid molecules and uses therefor are discussed further below.

In the event the nucleic acid molecules of the invention encode or act as antisense molecules, they can be used for example, to regulate translation of Tango-78, Tango-79, or Tango-81 mRNA.

The invention also encompasses nucleic acid molecules that hybridize under stringent conditions to a nucleic acid molecule encoding a Tango-78, Tango-79, or Tango-81 polypeptide (e.g., the protein encoding portion of SEQ ID NO:1, SEQ ID:3, or SEQ ID NO:5). The cDNA sequences described herein can be used to identify these nucleic acids, which include, for example, nucleic acids that encode homologous polypeptides in other species, and splice variants of the Tango-78, Tango-79, or Tango-81 gene in humans or other mammals. Accordingly, the invention features methods of detecting and isolating

these nucleic acid molecules. Using these methods, a sample (for example, a nucleic acid library, such as a cDNA or genomic library) is contacted (or "screened") with a Tango-78, Tango-79, or Tango-81-specific probe. The probe will selectively hybridize to nucleic acids encoding related polypeptides (or to complementary sequences thereof). The probe, which can contain at least 25 (for example, 25, 50, 100, or 200 nucleotides) can be produced using any of several standard methods (see, for example, Ausubel et al.,"Current Protocols in Molecular Biology, Vol. I," Green Publishing Associates, Inc., and John Wiley & Sons, Inc., NY, 1989). For example, the probe can be generated using PCR amplification methods in which oligonucleotide primers are used to amplify a Tango-78, Tango-79, or Tango-81-specific nucleic acid sequence that can be used as a probe to screen a nucleic acid library and thereby detect nucleic acid molecules (within the library) that hybridize to the probe.

One single-stranded nucleic acid is said to hybridize to another if a duplex forms between them. This occurs when one nucleic acid contains a sequence that is the reverse and complement of the other (this same arrangement gives rise to the natural interaction between the sense and antisense strands of DNA in the genome and underlies the configuration of the "double helix"). Complete complementarity between the hybridizing regions is not required in order for a duplex to form; it is only necessary that the number of paired bases is sufficient to maintain the duplex under the hybridization conditions used.

Typically, hybridization conditions are of low to moderate stringency. These conditions favor specific interactions between completely complementary sequences, but allow some non-specific interaction between less than

perfectly matched sequences to occur as well. After hybridization, the nucleic acids can be "washed" under moderate or high conditions of stringency to dissociate duplexes that are bound together by some non-specific interaction (the nucleic acids that form these duplexes are thus not completely complementary).

As is known in the art, the optimal conditions for washing are determined empirically, often by gradually increasing the stringency. The parameters that can be changed to affect stringency include, primarily, temperature and salt concentration. In general, the lower the salt concentration and the higher the temperature, the higher the stringency. Washing can be initiated at a low temperature (for example, room temperature) using a solution containing a salt concentration that is equivalent to or lower than that of the hybridization solution. Subsequent washing can be carried out using progressively warmer solutions having the same salt concentration. As alternatives, the salt concentration can be lowered and the temperature maintained in the washing step, or the salt concentration can be lowered and the temperature increased. Additional parameters can also be altered. For example, use of a destabilizing agent, such as formamide, alters the stringency conditions.

In reactions where nucleic acids are hybridized, the conditions used to achieve a given level of stringency will vary. There is not one set of conditions, for example, that will allow duplexes to form between all nucleic acids that are 85% identical to one another; hybridization also depends on unique features of each nucleic acid. The length of the sequence, the composition of the sequence (for example, the content of purine-like nucleotides versus the content of pyrimidine-like nucleotides) and the type of nucleic acid (for

example, DNA or RNA) affect hybridization. An additional consideration is whether one of the nucleic acids is immobilized (for example, on a filter).

An example of a progression from lower to higher stringency conditions is the following, where the salt content is given as the relative abundance of SSC (a salt solution containing sodium chloride and sodium citrate; 2X SSC is 10-fold more concentrated than 0.2X SSC). Nucleic acids are hybridized at 42°C in 2X SSC/0.1% SDS (sodium dodecylsulfate; a detergent) and then washed in 0.2X SSC/0.1% SDS at room temperature (for conditions of low stringency); 0.2X SSC/0.1% SDS at 42°C (for conditions of moderate stringency); and 0.1X SSC at 68°C (for conditions of high stringency). Washing can be carried out using only one of the conditions given, or each of the conditions can be used (for example, washing for 10-15 minutes each in the order listed above). Any or all of the washes can be repeated. As mentioned above, optimal conditions will vary and can be determined empirically.

A second set of conditions that are considered "stringent conditions" are those in which hybridization is carried out at 50°C in Church buffer (7% SDS, 0.5% $NaHPO_4$, 1 M EDTA, 1% BSA) and washing is carried out at 50°C in 2X SSC.

Once detected, the nucleic acid molecules can be isolated by any of a number of standard techniques (see, for example, Sambrook et al., "Molecular Cloning, A Laboratory Manual," 2nd Ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

The invention also encompasses: (a) expression vectors that contain any of the foregoing Tango-78, Tango-79, and Tango-81-related coding sequences and/or their complements (that is, "antisense" sequence); (b) expression vectors that contain any of the foregoing

Tango-78, Tango-79, or Tango-81-related coding sequences operatively associated with a regulatory element (examples of which are given below) that directs the expression of the coding sequences; (c) expression vectors containing, in addition to sequences encoding a Tango-78, Tango-79, or Tango-81 polypeptide, nucleic acid sequences that are unrelated to nucleic acid sequences encoding Tango-78, Tango-79, or Tango-81, such as molecules encoding a reporter or marker; and (d) genetically engineered host cells that contain any of the foregoing expression vectors and thereby express the nucleic acid molecules of the invention in the host cell.

Recombinant nucleic acid molecules can contain a sequence encoding a soluble Tango-78, Tango-79, or Tango-81 polypeptide; mature Tango-78, Tango-79, or Tango-81; or Tango-78, Tango-79, or Tango-81 having an added or endogenous signal sequence. A full length Tango-78, Tango-79, or Tango-81 polypeptide; a domain of Tango-78, Tango-79, or Tango-81; or a fragment thereof may be fused to additional polypeptides, as described below. Similarly, the nucleic acid molecules of the invention can encode the mature form of Tango-78, Tango-79, or Tango-81 or a form that encodes a polypeptide which facilitates secretion. In the latter instance, the polypeptide is typically referred to as a proprotein, which can be converted into an active form by removal of the signal sequence, for example, within the host cell. Proproteins can be converted into the active form of the protein by removal of the inactivating sequence.

The regulatory elements referred to above include, but are not limited to, inducible and non-inducible promoters, enhancers, operators and other elements, which are known to those skilled in the art, and which drive or otherwise regulate gene expression. Such regulatory elements include but are not limited to the

cytomegalovirus hCMV immediate early gene, the early or late promoters of SV40 adenovirus, the lac system, the trp system, the TAC system, the TRC system, the major operator and promoter regions of phage A, the control regions of fd coat protein, the promoter for 3-phosphoglycerate kinase, the promoters of acid phosphatase, and the promoters of the yeast α-mating factors.

Similarly, the nucleic acid can form part of a hybrid gene encoding additional polypeptide sequences, for example, sequences that function as a marker or reporter. Examples of marker or reporter genes include β-lactamase, chloramphenicol acetyltransferase (CAT), adenosine deaminase (ADA), aminoglycoside phosphotransferase (neo^r, $G418^r$), dihydrofolate reductase (DHFR), hygromycin-B-phosphotransferase (HPH), thymidine kinase (TK), lacZ (encoding β-galactosidase), and xanthine guanine phosphoribosyltransferase (XGPRT). As with many of the standard procedures associated with the practice of the invention, skilled artisans will be aware of additional useful reagents, for example, of additional sequences that can serve the function of a marker or reporter. Generally, the hybrid polypeptide will include a first portion and a second portion; the first portion being a Tango-78, Tango-79, or Tango-81 polypeptide and the second portion being, for example, the reporter described above or an immunoglobulin constant region.

The expression systems that may be used for purposes of the invention include, but are not limited to, microorganisms such as bacteria (for example, *E. coli* and *B. subtilis*) transformed with recombinant bacteriophage DNA, plasmid DNA, or cosmid DNA expression vectors containing the nucleic acid molecules of the invention; yeast (for example, *Saccharomyces* and *Pichia*) transformed with recombinant yeast expression vectors

containing the nucleic acid molecules of the invention; insect cell systems infected with recombinant virus expression vectors (for example, baculovirus) containing the nucleic acid molecules of the invention; plant cell systems infected with recombinant virus expression vectors (for example, cauliflower mosaic virus (CaMV) and tobacco mosaic virus (TMV)) or transformed with recombinant plasmid expression vectors (for example, Ti plasmid) containing Tango-78, Tango-79, or Tango-81 nucleotide sequences; or mammalian cell systems (for example, COS, CHO, BHK, 293, VERO, HeLa, MDCK, WI38, and NIH 3T3 cells) harboring recombinant expression constructs containing promoters derived from the genome of mammalian cells (for example, the metallothionein promoter) or from mammalian viruses (for example, the adenovirus late promoter and the vaccinia virus 7.5K promoter).

In bacterial systems, a number of expression vectors may be advantageously selected depending upon the use intended for the gene product being expressed. For example, when a large quantity of such a protein is to be produced, for the generation of pharmaceutical compositions containing Tango-78, Tango-79, or Tango-81 polypeptides or for raising antibodies to those polypeptides, vectors that are capable of directing the expression of high levels of fusion protein products that are readily purified may be desirable. Such vectors include, but are not limited to, the *E. coli* expression vector pUR278 (Ruther et al., *EMBO J.* 2:1791, 1983), in which the coding sequence of the insert may be ligated individually into the vector in frame with the lacZ coding region so that a fusion protein is produced; pIN vectors (Inouye and Inouye, *Nucleic Acids Res.* 13:3101-3109, 1985; Van Heeke and Schuster, *J. Biol. Chem.* 264:5503-5509, 1989); and the like. pGEX vectors may

also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. The pGEX vectors are designed to include thrombin or factor Xa protease cleavage sites so that the cloned target gene product can be released from the GST moiety.

In an insect system, *Autographa californica* nuclear polyhidrosis virus (AcNPV) can be used as a vector to express foreign genes. The virus grows in *Spodoptera frugiperda* cells. The coding sequence of the insert may be cloned individually into non-essential regions (for example the polyhedrin gene) of the virus and placed under control of an AcNPV promoter (for example the polyhedrin promoter). Successful insertion of the coding sequence will result in inactivation of the polyhedrin gene and production of non-occluded recombinant virus (i.e., virus lacking the proteinaceous coat coded for by the polyhedrin gene). These recombinant viruses are then used to infect *Spodoptera frugiperda* cells in which the inserted gene is expressed. (for example, see Smith et al., *J. Virol.* 46:584, 1983; Smith, U.S. Patent No. 4,215,051).

In mammalian host cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, the nucleic acid molecule of the invention may be ligated to an adenovirus transcription/translation control complex, for example, the late promoter and tripartite leader sequence. This chimeric gene may then be inserted in the adenovirus genome by *in vitro* or *in vivo* recombination. Insertion in a non-essential region of the viral genome (for example, region E1 or E3) will result in a

recombinant virus that is viable and capable of expressing a Tango-78, Tango-79, or Tango-81 gene product in infected hosts (for example, see Logan and Shenk, *Proc. Natl. Acad. Sci. USA* 81:3655-3659, 1984). Specific initiation signals may also be required for efficient translation of inserted nucleic acid molecules. These signals include the ATG initiation codon and adjacent sequences. In cases where an entire gene or cDNA, including its own initiation codon and adjacent sequences, is inserted into the appropriate expression vector, no additional translational control signals may be needed. However, in cases where only a portion of the coding sequence is inserted, exogenous translational control signals, including, perhaps, the ATG initiation codon, must be provided. Furthermore, the initiation codon must be in phase with the reading frame of the desired coding sequence to ensure translation of the entire insert. These exogenous translational control signals and initiation codons can be of a variety of origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of appropriate transcription enhancer elements, transcription terminators, etc. (see Bittner et al., *Methods in Enzymol.* 153:516-544, 1987).

In addition, a host cell strain may be chosen which modulates the expression of the inserted sequences, or modifies and processes the gene product in the specific fashion desired. Such modifications (for example, glycosylation) and processing (for example, cleavage) of protein products may be important for the function of the protein. Different host cells have characteristic and specific mechanisms for the post-translational processing and modification of proteins and gene products. Appropriate cell lines or host systems can be chosen to ensure the correct modification and

processing of the foreign protein expressed. To this end, eukaryotic host cells which possess the cellular machinery for proper processing of the primary transcript, glycosylation, and phosphorylation of the gene product may be used. The mammalian cell types listed above are among those that could serve as suitable host cells.

For long-term, high-yield production of recombinant proteins, stable expression is preferred. For example, cell lines which stably express the Tango-78, Tango-79, or Tango-81 sequences described above may be engineered. Rather than using expression vectors which contain viral origins of replication, host cells can be transformed with DNA controlled by appropriate expression control elements (for example, promoter, enhancer sequences, transcription terminators, polyadenylation sites, etc.), and a selectable marker. Following the introduction of the foreign DNA, engineered cells may be allowed to grow for 1-2 days in an enriched media, and then switched to a selective media. The selectable marker in the recombinant plasmid confers resistance to the selection and allows cells to stably integrate the plasmid into their chromosomes and grow to form foci which in turn can be cloned and expanded into cell lines. This method can advantageously be used to engineer cell lines which express Tango-78, Tango-79, or Tango-81. Such engineered cell lines may be particularly useful in screening and evaluation of compounds that affect the endogenous activity of the gene product.

A number of selection systems can be used. For example, the herpes simplex virus thymidine kinase (Wigler, et al., *Cell* 11:223, 1977), hypoxanthine-guanine phosphoribosyltransferase (Szybalska and Szybalski, *Proc. Natl. Acad. Sci. USA* 48:2026, 1962), and adenine phosphoribosyltransferase (Lowy, et al., *Cell* 22:817,

1980) genes can be employed in tk$^-$, hgprt$^-$ or aprt$^-$ cells, respectively. Also, anti-metabolite resistance can be used as the basis of selection for the following genes: dhfr, which confers resistance to methotrexate (Wigler et al., *Proc. Natl. Acad. Sci. USA* 77:3567, 1980; O'Hare et al., *Proc. Natl. Acad. Sci. USA* 78:1527, 1981); gpt, which confers resistance to mycophenolic acid (Mulligan and Berg, *Proc. Natl. Acad. Sci. USA* 78:2072, 1981); neo, which confers resistance to the aminoglycoside G-418 (Colberre-Garapin et al., *J. Mol. Biol.* 150:1, 1981); and hygro, which confers resistance to hygromycin (Santerre et al., *Gene* 30:147, 1984).

The nucleic acid molecules of the invention are useful for diagnosis of disorders associated with aberrant expression of Tango-78, Tango-79, or Tango-81. Tango-78, Tango-79, and Tango-81 nucleic acid molecules are also useful in genetic mapping and chromosome identification.

Tango-78, Tango-79, and Tango-81 Polypeptides

The Tango-78, Tango-79, and Tango-81 polypeptides described herein are those encoded by any of the nucleic acid molecules described above and include Tango-78, Tango-79, and Tango-81 fragments, mutants, truncated forms, and fusion proteins. These polypeptides can be prepared for a variety of uses, including but not limited to the generation of antibodies, as reagents in diagnostic assays, for the identification of other cellular gene products or compounds that can modulate the activity or expression of Tango-78, Tango-79, or Tango-81, and as pharmaceutical reagents useful for the treatment of disorders associated with aberrant expression or activity of Tango-78, Tango-79, or Tango-81.

Preferred polypeptides are substantially pure Tango-78, Tango-79, and Tango-81 polypeptides, including

those that correspond to the polypeptide with an intact signal sequence, and the secreted form of the polypeptide.

The invention also encompasses polypeptides that are functionally equivalent to Tango-78, Tango-79, or Tango-81. These polypeptides are equivalent to Tango-78, Tango-79, or Tango-81 in that they are capable of carrying out one or more of the functions of Tango-78, Tango-79, or Tango-81 in a biological system. Preferred Tango-78, Tango-79, or Tango-81 polypeptides have 20%, 40%, 50%, 75%, 80%, or even 90% of one or more of the biological activities of the full-length, mature human form of Tango-78, Tango-79, and Tango-81. Such comparisons are generally based on an assay of biological activity in which equal concentrations of the polypeptides are used and compared. The comparison can also be based on the amount of the polypeptide required to reach 50% of the maximal stimulation obtainable.

Functionally equivalent proteins can be those, for example, that contain additional or substituted amino acid residues. Substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues involved. Amino acids that are typically considered to provide a conservative substitution for one another are specified in the summary of the invention.

Polypeptides that are functionally equivalent to Tango-78, Tango-79, or Tango-81 can be made using random mutagenesis techniques well known to those skilled in the art. It is more likely, however, that such polypeptides will be generated by site-directed mutagenesis (again using techniques well known to those skilled in the art). These polypeptides may have increased functionality or decreased functionality.

To design functionally equivalent polypeptides, it is useful to distinguish between conserved positions and variable positions. This can be done by aligning the amino acid sequence of a protein of the invention from one species with its homolog from another species. Skilled artisans will recognize that conserved amino acid residues are more likely to be necessary for preservation of function. Thus, it is preferable that conserved residues are not altered.

Mutations within the coding sequence of nucleic acid molecules of the invention can be made to generate variant genes that are better suited for expression in a selected host cell. For example, N-linked glycosylation sites can be altered or eliminated to achieve, for example, expression of a homogeneous product that is more easily recovered and purified from yeast hosts which are known to hyperglycosylate N-linked sites. To this end, a variety of amino acid substitutions at one or both of the first or third amino acid positions of any one or more of the glycosylation recognition sequences which occur, and/or an amino acid deletion at the second position of any one or more of such recognition sequences, will prevent glycosylation at the modified tripeptide sequence (see, for example, Miyajima et al., *EMBO J.* 5:1193, 1986).

The polypeptides of the invention can be expressed fused to another polypeptide, for example, a marker polypeptide or fusion partner. For example, the polypeptide can be fused to a hexa-histidine tag to facilitate purification of bacterially expressed protein or a hemagglutinin tag to facilitate purification of protein expressed in eukaryotic cells.

A fusion protein may be readily purified by utilizing an antibody specific for the fusion protein being expressed. For example, a system described by

Janknecht et al. allows for the ready purification of non-denatured fusion proteins expressed in human cell lines (*Proc. Natl. Acad. Sci. USA* 88: 8972-8976, 1991). In this system, the gene of interest is subcloned into a vaccinia recombination plasmid such that the gene's open reading frame is translationally fused to an amino-terminal tag consisting of six histidine residues. Extracts from cells infected with recombinant vaccinia virus are loaded onto Ni^{2+}·nitriloacetic acid-agarose columns and histidine-tagged proteins are selectively eluted with imidazole-containing buffers.

The polypeptides of the invention can be chemically synthesized (for example, see Creighton, "Proteins: Structures and Molecular Principles," W.H. Freeman & Co., NY, 1983), or, perhaps more advantageously, produced by recombinant DNA technology as described herein. For additional guidance, skilled artisans may consult Ausubel et al. (*supra*), Sambrook et al. ("Molecular Cloning, A Laboratory Manual," Cold Spring Harbor Press, Cold Spring Harbor, NY, 1989), and, particularly for examples of chemical synthesis Gait, M.J. Ed. ("Oligonucleotide Synthesis," IRL Press, Oxford, 1984).

The invention also features polypeptides that interact with Tango-78, Tango-79, or Tango-81 (and the genes that encode them) and thereby alter the function of Tango-78, Tango-79, or Tango-81. Interacting polypeptides can be identified using methods known to those skilled in the art. One suitable method is the "two-hybrid system," which detects protein interactions *in vivo* (Chien et al., *Proc. Natl. Acad. Sci. USA*, 88:9578, 1991). A kit for practicing this method is available from Clontech (Palo Alto, CA).

Transgenic animals

Tango-78, Tango-79, and Tango-81 polypeptides can also be expressed in transgenic animals. These animals represent a model system for the study of disorders that are caused by or exacerbated by overexpression or underexpression of Tango-78, Tango-79, or Tango-81, and for the development of therapeutic agents that modulate the expression or activity of Tango-78, Tango-79, or Tango-81.

Transgenic animals can be farm animals (pigs, goats, sheep, cows, horses, rabbits, and the like) rodents (such as rats, guinea pigs, and mice), non-human primates (for example, baboons, monkeys, and chimpanzees), and domestic animals (for example, dogs and cats). Transgenic mice are especially preferred.

Any technique known in the art can be used to introduce a Tango-78, Tango-79, or Tango-81 transgene into animals to produce the founder lines of transgenic animals. Such techniques include, but are not limited to, pronuclear microinjection (U.S. Pat. No. 4,873,191); retrovirus mediated gene transfer into germ lines (Van der Putten et al., *Proc. Natl. Acad. Sci., USA* 82:6148, 1985); gene targeting into embryonic stem cells (Thompson et al., *Cell* 56:313, 1989); and electroporation of embryos (Lo, *Mol. Cell. Biol.* 3:1803, 1983).

The present invention provides for transgenic animals that carry a Tango-78, Tango-79, or Tango-81 transgene in all their cells, as well as animals that carry a transgene in some, but not all of their cells. That is, the invention provides for mosaic animals. The transgene can be integrated as a single transgene or in concatamers, e.g., head-to-head tandems or head-to-tail tandems. The transgene can also be selectively introduced into and activated in a particular cell type (Lasko et al., *Proc. Natl. Acad. Sci. USA* 89:6232, 1992).

The regulatory sequences required for such a cell-type specific activation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

When it is desired that the Tango-78, Tango-79, or Tango-81 transgene be integrated into the chromosomal site of the endogenous gene, gene targeting is preferred. Briefly, when such a technique is to be used, vectors containing some nucleotide sequences homologous to an endogenous Tango-78, Tango-79, or Tango-81 gene are designed for the purpose of integrating, via homologous recombination with chromosomal sequences, into and disrupting the function of the nucleotide sequence of the endogenous gene. The transgene also can be selectively introduced into a particular cell type, thus inactivating the endogenous Tango-78, Tango-79, or Tango-81 gene in only that cell type (Gu et al., *Science* 265:103, 1984). The regulatory sequences required for such a cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art. These techniques are useful for preparing "knock outs" lacking a functional gene.

Once transgenic animals have been generated, the expression of the recombinant Tango-78, Tango-79, or Tango-81 gene can be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to determine whether integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also be assessed using techniques which include, but are not limited to, Northern blot analysis of tissue samples obtained from the animal, *in situ* hybridization analysis, and RT-PCR. Biological samples can also be evaluated

immunocytochemically using antibodies specific for the Tango-78, Tango-79, or Tango-81 transgene product.

For a review of techniques that can be used to generate and assess transgenic animals, skilled artisans can consult Gordon (*Intl. Rev. Cytol.* 115:171-229, 1989), and may obtain additional guidance from, for example: Hogan et al. "Manipulating the Mouse Embryo" (Cold Spring Harbor Press, Cold Spring Harbor, NY, 1986; Krimpenfort et al., *Bio/Technology* 9:86, 1991; Palmiter et al., *Cell* 41:343, 1985; Kraemer et al., "Genetic Manipulation of the Early Mammalian Embryo," Cold Spring Harbor Press, Cold Spring Harbor, NY, 1985; Hammer et al., *Nature* 315:680, 1985; Purcel et al., *Science*, 244:1281, 1986; Wagner et al., U.S. Patent No. 5,175,385; and Krimpenfort et al., U.S. Patent No. 5,175,384 (the latter two publications are hereby incorporated by reference).

Anti-Tango-78, Tango-79, or Tango-81 Antibodies

Human Tango-78, Tango-79, and Tango-81 polypeptides (or immunogenic fragments or analogs) can be used to raise antibodies useful in the invention; such polypeptides can be produced by recombinant techniques or synthesized (see, for example, "Solid Phase Peptide Synthesis," *supra*; Ausubel et al., *supra*). In general, the peptides can be coupled to a carrier protein, such as KLH, as described in Ausubel et al., *supra*, mixed with an adjuvant, and injected into a host mammal. Antibodies can be purified by peptide antigen affinity chromatography.

In particular, various host animals can be immunized by injection with a Tango-78, Tango-79, or Tango-81 polypeptide. Host animals include rabbits, mice, guinea pigs, and rats. Various adjuvants that can be used to increase the immunological response depend on the host species and include Freund's adjuvant (complete and incomplete), mineral gels such as aluminum hydroxide,

surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, and dinitrophenol. Potentially useful human adjuvants include BCG (bacille Calmette-Guerin) and *Corynebacterium parvum*. Polyclonal antibodies are heterogeneous populations of antibody molecules that are contained in the sera of the immunized animals.

Antibodies within the invention therefore include polyclonal antibodies and, in addition, monoclonal antibodies, humanized or chimeric antibodies, single chain antibodies, Fab fragments, $F(ab')_2$ fragments, and molecules produced using a Fab expression library.

Monoclonal antibodies, which are homogeneous populations of antibodies to a particular antigen, can be prepared using the Tango-78, Tango-79, or Tango-81 polypeptides described above and standard hybridoma technology (see, for example, Kohler et al., *Nature* 256:495, 1975; Kohler et al., *Eur. J. Immunol.* 6:511, 1976; Kohler et al., *Eur. J. Immunol.* 6:292, 1976; Hammerling et al., "Monoclonal Antibodies and T Cell Hybridomas," Elsevier, NY, 1981; Ausubel et al., *supra*).

In particular, monoclonal antibodies can be obtained by any technique that provides for the production of antibody molecules by continuous cell lines in culture such as described in Kohler et al., *Nature* 256:495, 1975, and U.S. Patent No. 4,376,110; the human B-cell hybridoma technique (Kosbor et al., *Immunology Today* 4:72, 1983; Cole et al., *Proc. Natl. Acad. Sci. USA* 80:2026, 1983), and the EBV-hybridoma technique (Cole et al., "Monoclonal Antibodies and Cancer Therapy," Alan R. Liss, Inc., pp. 77-96, 1983). Such antibodies can be of any immunoglobulin class including IgG, IgM, IgE, IgA, IgD and any subclass thereof. The hybridoma producing the mAb of this invention may be cultivated *in vitro* or *in vivo*. The ability to produce high titers of mAbs

in vivo makes this a particularly useful method of production.

Once produced, polyclonal or monoclonal antibodies are tested for specific Tango-78, Tango-79, or Tango-81 recognition by Western blot or immunoprecipitation analysis by standard methods, e.g., as described in Ausubel et al., *supra*. Antibodies that specifically recognize and bind to Tango-78, Tango-79, or Tango-81 are useful in the invention. For example, such antibodies can be used in an immunoassay to monitor the level of Tango-78, Tango-79, or Tango-81 produced by a mammal (for example, to determine the amount or subcellular location of Tango-78, Tango-79, or Tango-81).

Preferably, antibodies of the invention are produced using fragments of the Tango-78, Tango-79, or Tango-81 protein which lie outside highly conserved regions and appear likely to be antigenic, by criteria such as high frequency of charged residues. In one specific example, such fragments are generated by standard techniques of PCR, and are then cloned into the pGEX expression vector (Ausubel et al., *supra*). Fusion proteins are expressed in *E. coli* and purified using a glutathione agarose affinity matrix as described in Ausubel, et al., *supra*.

In some cases it may be desirable to minimize the potential problems of low affinity or specificity of antisera. In such circumstances, two or three fusions can be generated for each protein, and each fusion can be injected into at least two rabbits. Antisera can be raised by injections in a series, preferably including at least three booster injections.

Antisera may also checked for its ability to immunoprecipitate recombinant Tango-78, Tango-79, and Tango-81 proteins or control proteins, such as glucocorticoid receptor, CAT, or luciferase.

The antibodies can be used, for example, in the detection of the Tango-78, Tango-79, or Tango-81 in a biological sample as part of a diagnostic assay. Antibodies also can be used in a screening assay to measure the effect of a candidate compound on expression or localization of Tango-78, Tango-79, or Tango-81. Additionally, such antibodies can be used in conjunction with the gene therapy techniques described to, for example, evaluate normal and/or genetically engineered Tango-78, Tango-79, and Tango-81-expressing cells prior to their introduction into the patient.

In addition, techniques developed for the production of "chimeric antibodies" (Morrison et al., *Proc. Natl. Acad. Sci. USA*, 81:6851, 1984; Neuberger et al., *Nature*, 312:604, 1984; Takeda et al., *Nature*, 314:452, 1984) by splicing the genes from a mouse antibody molecule of appropriate antigen specificity together with genes from a human antibody molecule of appropriate biological activity can be used. A chimeric antibody is a molecule in which different portions are derived from different animal species, such as those having a variable region derived from a murine mAb and a human immunoglobulin constant region.

Alternatively, techniques described for the production of single chain antibodies (U.S. Patent Nos. 4,946,778, 4,946,778, and 4,704,692) can be adapted to produce single chain antibodies against a Tango-78, Tango-79, or Tango-81 or polypeptide. Single chain antibodies are formed by linking the heavy and light chain fragments of the Fv region via an amino acid bridge, resulting in a single chain polypeptide.

Antibody fragments that recognize and bind to specific epitopes can be generated by known techniques. For example, such fragments include but are not limited to $F(ab')_2$ fragments that can be produced by pepsin

digestion of the antibody molecule, and Fab fragments that can be generated by reducing the disulfide bridges of $F(ab')_2$ fragments. Alternatively, Fab expression libraries can be constructed (Huse et al., *Science*, 246:1275, 1989) to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity.

Antibodies to Tango-78, Tango-79, or Tango-81 can, in turn, be used to generate anti-idiotype antibodies that resemble a portion of the protein using techniques well known to those skilled in the art (see, e.g., Greenspan et al., *FASEB J.* 7:437, 1993; Nissinoff, *J. Immunol.* 147:2429, 1991). For example, antibodies that bind to the protein and competitively inhibit the binding of a binding partner of the protein can be used to generate anti-idiotypes that resemble a binding partner binding domain of the protein and, therefore, bind and neutralize a binding partner of the protein. Such neutralizing anti-idiotypic antibodies or Fab fragments of such anti-idiotypic antibodies can be used in therapeutic regimens.

Antibodies can be humanized by methods known in the art. For example, monoclonal antibodies with a desired binding specificity can be commercially humanized (Scotgene, Scotland; Oxford Molecular, Palo Alto, CA). Fully human antibodies, such as those expressed in transgenic animals are also features of the invention (Green et al., *Nature Genetics* 7:13-21, 1994; see also U.S. Patents 5,545,806 and 5,569,825, both of which are hereby incorporated by reference).

The methods described herein in which anti-Tango-78, Tango-79, or Tango-81 antibodies are employed may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one specific Tango-78, Tango-79, or Tango-81 antibody reagent described herein, which may be conveniently used, for example, in

clinical settings, to diagnose patients exhibiting symptoms disorders associated with abberent expression of Tango-78, Tango-79, or Tango-81.

Antisense Nucleic Acids

Treatment regimes based on an "antisense" approach involve the design of oligonucleotides (either DNA or RNA) that are complementary to Tango-78, Tango-79, or Tango-81 mRNA. These oligonucleotides bind to the complementary Tango-78, Tango-79, or Tango-81 mRNA transcripts and prevent translation. Absolute complementarity, although preferred, is not required. A sequence "complementary" to a portion of an RNA, as referred to herein, means a sequence having sufficient complementarily to be able to hybridize with the RNA, forming a stable duplex; in the case of double-stranded antisense nucleic acids, a single strand of the duplex DNA may be tested, or triplex formation may be assayed. The ability to hybridize will depend on both the degree of complementarily and the length of the antisense nucleic acid. Generally, the longer the hybridizing nucleic acid, the more base mismatches with an RNA it may contain and still form a stable duplex (or triplex, as the case may be). One skilled in the art can ascertain a tolerable degree of mismatch by use of standard procedures to determine the melting point of the hybridized complex.

Oligonucleotides that are complementary to the 5′ end of the message, e.g., the 5′ untranslated sequence up to and including the AUG initiation codon, should work most efficiently at inhibiting translation. However, sequences complementary to the 3′ untranslated sequences of mRNAs recently have been shown to be effective at inhibiting translation of mRNAs as well (Wagner, *Nature* 372:333, 1984). Thus, oligonucleotides complementary to either the 5′ or 3′ non-translated, non-coding regions of

the gene could be used in an antisense approach to inhibit translation of endogenous mRNA. Oligonucleotides complementary to the 5′ untranslated region of the mRNA should include the complement of the AUG start codon.

Antisense oligonucleotides complementary to mRNA coding regions are less efficient inhibitors of translation but could be used in accordance with the invention. Whether designed to hybridize to the 5′, 3′, or coding region of an mRNA, antisense nucleic acids should be at least six nucleotides in length, and are preferably oligonucleotides ranging from 6 to about 50 nucleotides in length. In specific aspects the oligonucleotide is at least 10 nucleotides, at least 17 nucleotides, at least 25 nucleotides, or at least 50 nucleotides.

Regardless of the choice of target sequence, it is preferred that *in vitro* studies are first performed to quantitate the ability of the antisense oligonucleotide to inhibit gene expression. It is preferred that these studies utilize controls that distinguish between antisense gene inhibition and nonspecific biological effects of oligonucleotides. It is also preferred that these studies compare levels of the target RNA or protein with that of an internal control RNA or protein. Additionally, it is envisioned that results obtained using the antisense oligonucleotide are compared with those obtained using a control oligonucleotide. It is preferred that the control oligonucleotide is of approximately the same length as the test oligonucleotide and that the nucleotide sequence of the oligonucleotide differs from the antisense sequence no more than is necessary to prevent specific hybridization to the target sequence.

The oligonucleotides can be DNA or RNA or chimeric mixtures or derivatives or modified versions thereof,

single-stranded or double-stranded. The oligonucleotide can be modified at the base moiety, sugar moiety, or phosphate backbone, for example, to improve stability of the molecule, hybridization, etc. The oligonucleotide may include other appended groups such as peptides (e.g., for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (as described, e.g., in Letsinger et al., *Proc. Natl. Acad. Sci. USA* 86:6553, 1989; Lemaitre et al., *Proc. Natl. Acad. Sci. USA* 84:648, 1987; PCT Publication No. WO 88/09810) or the blood-brain barrier (see, for example, PCT Publication No. WO 89/10134), or hybridization-triggered cleavage agents (see, for example, Krol et al., *BioTechniques* 6:958, 1988), or intercalating agents (see, for example, Zon, *Pharm. Res.* 5:539, 1988). To this end, the oligonucleotide can be conjugated to another molecule, e.g., a peptide, hybridization triggered cross-linking agent, transport agent, or hybridization-triggered cleavage agent.

The antisense oligonucleotide may comprise at least one modified base moiety which is selected from the group including, but not limited to, 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xantine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethyl-aminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic

acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-theouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 2-(3-amino-3-N-2-carboxypropl) uracil, (acp3)w, and 2,6-diaminopurine.

The antisense oligonucleotide may also comprise at least one modified sugar moiety selected from the group including, but not limited to, arabinose, 2-fluoroarabinose, xylulose, and hexose.

In yet another embodiment, the antisense oligonucleotide comprises at least one modified phosphate backbone selected from the group consisting of a phosphorothioate, a phosphorodithioate, a phosphoramidothioate, a phosphoramidate, a phosphordiamidate, a methylphosphonate, an alkyl phosphotriester, and a formacetal, or an analog of any of these backbones.

In yet another embodiment, the antisense oligonucleotide is an α-anomeric oligonucleotide. An α-anomeric oligonucleotide forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β-units, the strands run parallel to each other (Gautier et al., *Nucl. Acids. Res.* 15:6625, 1987). The oligonucleotide is a 2'-0-methylribonucleotide (Inoue et al., *Nucl. Acids Res.* 15:6131, 1987), or a chimeric RNA-DNA analog (Inoue et al., *FEBS Lett.* 215:327, 1987).

Antisense oligonucleotides of the invention can be synthesized by standard methods known in the art, e.g., by use of an automated DNA synthesizer (such as are commercially available from Biosearch, Applied Biosystems, etc.). As examples, phosphorothioate oligonucleotides can be synthesized by the method of Stein et al. (*Nucl. Acids Res.* 16:3209, 1988), and methylphosphonate oligonucleotides can be prepared by use

of controlled pore glass polymer supports (Sarin et al., *Proc. Natl. Acad. Sci. USA* 85:7448, 1988).

The antisense molecules should be delivered to cells that express Tango-78, Tango-79, or Tango-81 *in vivo*. A number of methods have been developed for delivering antisense DNA or RNA to cells; e.g., antisense molecules can be injected directly into the tissue site, or modified antisense molecules, designed to target the desired cells (e.g., antisense linked to peptides or antibodies that specifically bind receptors or antigens expressed on the target cell surface) can be administered systemically.

However, it is often difficult to achieve intracellular concentrations of the antisense molecule sufficient to suppress translation of endogenous mRNAs. Therefore, a preferred approach uses a recombinant DNA construct in which the antisense oligonucleotide is placed under the control of a strong *pol* III or *pol* II promoter. The use of such a construct to transfect target cells in the patient will result in the transcription of sufficient amounts of single stranded RNAs that will form complementary base pairs with the endogenous Tango-78, Tango-79, or Tango-81 transcripts and thereby prevent translation of the endogenous mRNA. For example, a vector can be introduced *in vivo* such that it is taken up by a cell and directs the transcription of an antisense RNA. Such a vector can remain episomal or become chromosomally integrated, as long as it can be transcribed to produce the desired antisense RNA.

Such vectors can be constructed by recombinant DNA technology methods standard in the art. Vectors can be plasmid, viral, or others known in the art, used for replication and expression in mammalian cells. Expression of the sequence encoding the antisense RNA can be by any promoter known in the art to act in mammalian,

preferably human cells. Such promoters can be inducible or constitutive. Such promoters include, but are not limited to: the SV40 early promoter region (Bernoist et al., *Nature* 290:304, 1981); the promoter contained in the 3' long terminal repeat of Rous sarcoma virus (Yamamoto et al., *Cell* 22:787-797, 1988); the herpes thymidine kinase promoter (Wagner et al., *Proc. Natl. Acad. Sci. USA* 78:1441, 1981); or the regulatory sequences of the metallothionein gene (Brinster et al., *Nature* 296:39, 1988).

Ribozymes

Ribozyme molecules designed to catalytically cleave Tango-78, Tango-79, or Tango-81 mRNA transcripts can be used to prevent translation of Tango-78, Tango-79, or Tango-81 mRNA. (see, e.g., PCT Publication WO 90/11364; Saraver et al., *Science* 247:1222, 1990). While various ribozymes that cleave mRNA at site-specific recognition sequences can be used to destroy Tango-78, Tango-79, or Tango-81 mRNAs, the use of hammerhead ribozymes is preferred. Hammerhead ribozymes cleave mRNAs at locations dictated by flanking regions that form complementary base pairs with the target mRNA. The sole requirement is that the target mRNA have the following sequence of two bases: 5'-UG-3'. The construction and production of hammerhead ribozymes is well known in the art (Haseloff et al., *Nature* 334:585, 1988). There are numerous examples of potential hammerhead ribozyme cleavage sites within the nucleotide sequence of human Tango-78, Tango-79, and Tango-81 cDNA. Preferably, the ribozyme is engineered so that the cleavage recognition site is located near the 5' end of the mRNA, i.e., to increase efficiency and minimize the intracellular accumulation of non-functional mRNA transcripts.

The ribozymes of the present invention also include RNA endoribonucleases (hereinafter "Cech-type

ribozymes"), such as the one that occurs naturally in *Tetrahymena Thermophila* (known as the IVS or L-19 IVS RNA), and which has been extensively described by Cech and his collaborators (Zaug et al., *Science* 224:574, 1984; Zaug et al., *Science*, 231:470, 1986; Zug et al., *Nature* 324:429, 1986; PCT Application No. WO 88/04300; and Been et al., *Cell* 47:207, 1986). The Cech-type ribozymes have an eight base-pair sequence that hybridizes to a target RNA sequence, whereafter cleavage of the target RNA takes place. The invention encompasses those Cech-type ribozymes that target eight base-pair active site sequences present in Tango-78, Tango-79, and Tango-81.

As in the antisense approach, the ribozymes can be composed of modified oligonucleotides (e.g., for improved stability, targeting, etc.), and should be delivered to cells which express Tango-78, Tango-79, or Tango-81 *in vivo*. A preferred method of delivery involves using a DNA construct "encoding" the ribozyme under the control of a strong constitutive *pol* III or *pol* II promoter, so that transfected cells will produce sufficient quantities of the ribozyme to destroy endogenous messages and inhibit translation. Because ribozymes, unlike antisense molecules, are catalytic, a lower intracellular concentration is required for efficiency.

Other Methods for Reducing Tango-78, Tango-79, and Tango-81 Expression

Endogenous Tango-78, Tango-79, and Tango-81 gene expression can also be reduced by inactivating the gene or its promoter using targeted homologous recombination (see, e.g., U.S. Patent No. 5,464,764). For example, a mutant, non-functional Tango-78, Tango-79, or Tango-81 gene (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous Tango-78, Tango-79, or Tango-81 gene (either the coding regions or regulatory

regions of the Tango-78, Tango-79, or Tango-81 gene) can be used, with or without a selectable marker and/or a negative selectable marker, to transfect cells that express Tango-78, Tango-79, or Tango-81 *in vivo*. Insertion of the DNA construct, via targeted homologous recombination, results in inactivation of the Tango-78, Tango-79, or Tango-81 gene. Such approaches are particularly suited for use in the agricultural field where modifications to ES (embryonic stem) cells can be used to generate animal offspring with an inactive Tango-78, Tango-79, or Tango-81. However, this approach can be adapted for use in humans, provided the recombinant DNA constructs are directly administered or targeted to the required site *in vivo* using appropriate viral vectors.

Alternatively, endogenous Tango-78, Tango-79, or Tango-81 gene expression can be reduced by targeting deoxyribonucleotide sequences complementary to the regulatory region of the Tango-78, Tango-79, or Tango-81 gene (*i.e.*, the Tango-78, Tango-79, or Tango-81 promoter and/or enhancers) to form triple helical structures that prevent transcription of the Tango-78, Tango-79, or Tango-81 gene in target cells in the body (Helene *Anticancer Drug Res.* 6:569, 1981; Helene et al., *Ann. N.Y. Acad. Sci.* 660:27, 1992; and Maher, *Bioassays* 14:807, 1992).

Detecting Proteins Associated with Tango-78, Tango-79, or Tango-81

The invention also features polypeptides which interact with Tango-78, Tango-79, or Tango-81. Any method suitable for detecting protein-protein interactions may be employed for identifying transmembrane proteins, intracellular, or extracellular proteins that interact with Tango-78, Tango-79, or Tango-81. Among the traditional methods which may be employed are co-immunoprecipitation, cross-linking and

co-purification through gradients or chromatographic columns of cell lysates or proteins obtained from cell lysates and the use of Tango-78, Tango-79, or Tango-81 to identify proteins in the lysate that interact with Tango-78, Tango-79, or Tango-81. For these assays, the Tango-78, Tango-79, or Tango-81 polypetide can be full length Tango-78, Tango-79, or Tango-81, a soluble extracellular domain of Tango-78, Tango-79, and Tango-81, or some other suitable Tango-78, Tango-79, or Tango-81 polypeptide. Once isolated, such an interacting protein can be identified and cloned and then used, in conjunction with standard techniques, to identify proteins with which it interacts. For example, at least a portion of the amino acid sequence of a protein which interacts with the Tango-78, Tango-79, or Tango-81 can be ascertained using techniques well known to those of skill in the art, such as via the Edman degradation technique. The amino acid sequence obtained may be used as a guide for the generation of oligonucleotide mixtures that can be used to screen for gene sequences encoding the interacting protein. Screening may be accomplished, for example, by standard hybridization or PCR techniques. Techniques for the generation of oligonucleotide mixtures and the screening are well-known. (Ausubel, *supra;* and "PCR Protocols: A Guide to Methods and Applications," Innis et al., eds. Academic Press, Inc., NY, 1990).

Additionally, methods may be employed which result directly in the identification of genes which encode proteins which interact with Tango-78, Tango-79, or Tango-81. These methods include, for example, screening expression libraries, in a manner similar to the well known technique of antibody probing of λgt11 libraries, using labeled Tango-78, Tango-79, or Tango-81 polypeptide or a Tango-78, Tango-79, or Tango-81 fusion protein, e.g., a Tango-78, Tango-79, or Tango-81 polypeptide or

domain fused to a marker such as an enzyme, fluorescent dye, a luminescent protein, or to an IgFc domain.

There are also methods which are capable of detecting protein interaction. A method which detects protein interactions *in vivo* is the two-hybrid system (Chien et al., *Proc. Natl. Acad. Sci. USA*, 88:9578, 1991). A kit for practicing this method is available from Clontech (Palo Alto, CA).

Identification of a Tango-78, Tango-79, or Tango-81 Receptor

Receptors of Tango-78, Tango-79, or Tango-81 can be identified as follows. First cells or tissues which bind Tango-78, Tango-79, or Tango-81 are identified. An expression library is prepared using mRNA isolated from Tango-78, Tango-79, or Tango-81 binding cells. The expression library is used to transfect; eulcaryatic cells, e.g., CHO cells. Detectably labelled Tango-78, Tango-79, or Tango-81 is used to identify clones which bind Tango-78, Tango-79, or Tango-81. These clones are isolated and purified. The expression plasmid is then isolated from the Tango-78, Tango-79, or Tango-81-binding clones. These expression plasmids will encode putative Tango-78, Tango-79, or Tango-81 receptors.

Cells or tissues bearing a Tango-78, Tango-79, or Tango-81 receptor can be identified by exposing detectably labelled Tango-78, Tango-79, or Tango-81 to various cells lines and tissues. Alternatively a microphysiometer can be used to determine whether a selected cells responds to the presence of a cell receptor ligand (McConnel et al., *Science* 257:1906, 1992).

Compounds which bind Tango-78, Tango-79, or Tango-81

Compounds which bind Tango-78, Tango-79, or Tango-81 can be identified using any standard binding assay. For example, candidate compounds can be bound to a solid support. Tango-78, Tango-79, or Tango-81 is then exposed to the immobilized compound and binding is measured (European Patent Application 84/03564).

Effective Dose

Toxicity and therapeutic efficacy of the polypeptides of the invention and the compounds that modulate their expression or activity can be determined by standard pharmaceutical procedures, using either cells in culture or experimental animals to determine the LD_{50} (the dose lethal to 50% of the population) and the ED_{50} (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD_{50}/ED_{50}. Polypeptides or other compounds that exhibit large therapeutic indices are preferred. While compounds that exhibit toxic side effects may be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED_{50} with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays.

A dose may be formulated in animal models to achieve a circulating plasma concentration range that includes the IC_{50} (that is, the concentration of the test compound which achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma may be measured, for example, by high performance liquid chromatography.

Formulations and Use

Pharmaceutical compositions for use in accordance with the present invention may be formulated in conventional manner using one or more physiologically acceptable carriers or excipients.

Thus, the compounds and their physiologically acceptable salts and solvates may be formulated for administration by inhalation or insufflation (either through the mouth or the nose) or oral, buccal, parenteral or rectal administration.

For oral administration, the pharmaceutical compositions may take the form of, for example, tablets or capsules prepared by conventional means with pharmaceutically acceptable excipients such as binding agents (for example, pregelatinised maize starch, polyvinylpyrrolidone or hydroxypropyl methylcellulose); fillers (for example, lactose, microcrystalline cellulose or calcium hydrogen phosphate); lubricants (for example, magnesium stearate, talc or silica); disintegrants (for example, potato starch or sodium starch glycolate); or wetting agents (for example, sodium lauryl sulphate). The tablets may be coated by methods well known in the art. Liquid preparations for oral administration may take the form of, for example, solutions, syrups or suspensions, or they may be presented as a dry product for constitution with water or other suitable vehicle before use. Such liquid preparations may be prepared by

conventional means with pharmaceutically acceptable additives such as suspending agents (for example, sorbitol syrup, cellulose derivatives or hydrogenated edible fats); emulsifying agents (for example, lecithin or acacia); non-aqueous vehicles (for example, almond oil, oily esters, ethyl alcohol or fractionated vegetable oils); and preservatives (for example, methyl or propyl-p-hydroxybenzoates or sorbic acid). The preparations may also contain buffer salts, flavoring, coloring and sweetening agents as appropriate. Preparations for oral administration may be suitably formulated to give controlled release of the active compound.

For buccal administration the compositions may take the form of tablets or lozenges formulated in conventional manner.

For administration by inhalation, the compounds for use according to the present invention are conveniently delivered in the form of an aerosol spray presentation from pressurized packs or a nebulizer, with the use of a suitable propellant, for example, dichlorodifluoromethane, trichlorofluoromethane, dichlorotetrafluoroethane, carbon dioxide or other suitable gas. In the case of a pressurized aerosol the dosage unit may be determined by providing a valve to deliver a metered amount. Capsules and cartridges of, for example, gelatin for use in an inhaler or insufflator may be formulated containing a powder mix of the compound and a suitable powder base such as lactose or starch.

The compounds may be formulated for parenteral administration by injection, for example, by bolus injection or continuous infusion. Formulations for injection may be presented in unit dosage form, for example, in ampoules or in multi-dose containers, with an added preservative. The compositions may take such forms as suspensions, solutions or emulsions in oily or aqueous

vehicles, and may contain formulatory agents such as suspending, stabilizing and/or dispersing agents. Alternatively, the active ingredient may be in powder form for constitution with a suitable vehicle, for example, sterile pyrogen-free water, before use.

The compounds may also be formulated in rectal compositions such as suppositories or retention enemas, for example, containing conventional suppository bases such as cocoa butter or other glycerides.

In addition to the formulations described previously, the compounds may also be formulated as a depot preparation. Such long acting formulations may be administered by implantation (for example subcutaneously or intramuscularly) or by intramuscular injection. Thus, for example, the compounds may be formulated with suitable polymeric or hydrophobic materials (for example as an emulsion in an acceptable oil) or ion exchange resins, or as sparingly soluble derivatives, for example, as a sparingly soluble salt.

The compositions may, if desired, be presented in a pack or dispenser device which may contain one or more unit dosage forms containing the active ingredient. The pack may for example comprise metal or plastic foil, such as a blister pack. The pack or dispenser device may be accompanied by instructions for administration.

The therapeutic compositions of the invention can also contain a carrier or excipient, many of which are known to skilled artisans. Excipients which can be used include buffers (for example, citrate buffer, phosphate buffer, acetate buffer, and bicarbonate buffer), amino acids, urea, alcohols, ascorbic acid, phospholipids, proteins (for example, serum albumin), EDTA, sodium chloride, liposomes, mannitol, sorbitol, and glycerol. The nucleic acids, polypeptides, antibodies, or modulatory compounds of the invention can be administered

by any standard route of administration. For example, administration can be parenteral, intravenous, subcutaneous, intramuscular, intracranial, intraorbital, opthalmic, intraventricular, intracapsular, intraspinal, intracisternal, intraperitoneal, transmucosal, or oral. The modulatory compound can be formulated in various ways, according to the corresponding route of administration. For example, liquid solutions can be made for ingestion or injection; gels or powders can be made for ingestion, inhalation, or topical application. Methods for making such formulations are well known and can be found in, for example, "Remington's Pharmaceutical Sciences." It is expected that the preferred route of administration will be intravenous.

Examples

Tango-78 cDNA (SEQ ID NO:1; FIG. 1) was isolated from a human bone marrow cDNA library (Clonetech; Palo Alto, CA). This Tango-78 cDNA encodes a 169 amino acid portion of Tango-78, a novel protein (SEQ ID NO:2; FIG. 1) that is highly homologous to the murine nodal protein (Collignon et al., *Nature* 381:155, 1996).

The Tango-78 cDNA (SEQ ID NO:1; FIG. 1) described herein was isolated using the method described in U.S. Serial No. 08/752,307 (filed November 19, 1996), hereby incorporated by reference. Tango-78 protein (SEQ ID NO:2; FIG. 1) is highly homologous to murine nodal protein (Collignon et al., *supra*; FIG.

Tango-79 cDNA (SEQ ID NO:3; FIG. 2) was isolated from a human fetal brain library (Clonetech; Palo Alto, CA). This Tango-78 cDNA encodes a 615 amino acid protein (SEQ ID NO:4; FIG. 2) that is homologous to *Drosophila Melanogaster* slit protein (Taguchi et al., *Brain Res. Mol. Brain Res.* 35:31, 1996).

The Tango-79 cDNA (SEQ ID NO:3; FIG. 2) described herein was isolated using the method described in U.S. Serial No. 08/752,307 (filed November 19, 1996), hereby incorporated by reference. Tango-79 protein (SEQ ID NO:4; FIG. 2) is homologous to D45913 (leucine rich repeat protein) (FIG. 5). Northern blot analysis of Tango-79 mRNA show that an approximate 3.0 kB and an approximate 3.5 kB transpcript are expressed in the brain. Tango-79 function can be studied by overexpressing the protein in mouse brain.

Tango-81 cDNA was isolated from a human fetal brain library. This Tango-81 cDNA (SEQ ID NO:5; FIG. 3) encodes a 261 amino acid protein (SEQ ID NO:6; FIG. 3).

The Tango-81 cDNA described herein was isolated using the method described in U.S. Serial No. 08/752,307 (filed November 19, 1996), hereby incorporated by reference. Northern analysis of Tango-81 expression reveals that it is expressed in heart, brain, spleen, lung, liver, skeletal muscle, kidneys and testis (FIG. 6).

What is claimed is:

1. An isolated nucleic acid molecule selected from the group consisting of:

a) a nucleic acid molecule comprising a nucleotide sequence which is at least 55% identical to the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, the cDNA insert of the plasmid deposited with ATCC as Accession Number _____, the cDNA insert of the plasmid deposited with ATCC as Accession Number _____, the cDNA insert of the plasmid deposited with ATCC as Accession Number _____or a complement thereof;

b) a nucleic acid molecule comprising a fragment of at least 300 nucleotides of the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, the cDNA insert of the plasmid deposited with ATCC as Accession Number _____ , the cDNA insert of the plasmid deposited with ATCC as Accession Number _____, the cDNA insert of the plasmid deposited with ATCC as Accession Number _____ or a complement thereof;

c) a nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number _____, an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number _____, or an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number _____;

d) a nucleic acid molecule which encodes a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO: 4, SEQ ID NO:6, wherein the fragment comprises at least 15 contiguous amino acids of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, the polypeptide encoded by the cDNA insert of the plasmid

deposited with ATCC as Accession Number _____, the polypeptide encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number _____, or the polypeptide encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number _____; and

e) a nucleic acid molecule which encodes a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number _____, an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number _____, or an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number _____, wherein the nucleic acid molecule hybridizes to a nucleic acid molecule comprising SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, or a complement thereof under stringent conditions.

2. The isolated nucleic acid molecule of claim 1, which is selected from the group consisting of:

a) a nucleic acid comprising the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, the cDNA insert of the plasmid deposited with ATCC as Accession Number _____, the cDNA insert of the plasmid deposited with ATCC as Accession Number _____, the cDNA insert of the plasmid deposited with ATCC as Accession Number _____, or a complement thereof; and

b) a nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number _____, an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number _____, or

an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number _____.

3. The nucleic acid molecule of claim 1 further comprising vector nucleic acid sequences.

4. The nucleic acid molecule of claim 1 further comprising nucleic acid sequences encoding a heterologous polypeptide.

5. A host cell which contains the nucleic acid molecule of claim 1.

6. The host cell of claim 5 which is a mammalian host cell.

7. A non-human mammalian host cell containing the nucleic acid molecule of claim 1.

8. An isolated polypeptide selected from the group consisting of:

a) a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:6, wherein the fragment comprises at least 15 contiguous amino acids of SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:6;

b) a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number _____, an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number _____, or an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number _____,

wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to a nucleic acid molecule comprising SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, or a complement thereof under stringent conditions; and

c) a polypeptide which is encoded by a nucleic acid molecule comprising a nucleotide sequence which is at least 55% identical to a nucleic acid comprising the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, or a complement thereof.

9. The isolated polypeptide of claim 8 comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number ____, an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number _____, or an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number _____.

10. The polypeptide of claim 8 further comprising heterologous amino acid sequences.

11. An antibody which selectively binds to a polypeptide of claim 8.

12. A method for producing a polypeptide selected from the group consisting of:

a) a polypeptide comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number _____, an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number _____,

or an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number _____;

b) a polypeptide comprising a fragment of the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number ____, an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number _____, or an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number _____, wherein the fragment comprises at least 15 contiguous amino acids of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number _____, an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number _____, or an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number _____; and

c) a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number _____, an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number _____, or an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number _____, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to a nucleic acid molecule comprising SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, or a complement thereof under stringent conditions;

comprising culturing the host cell of claim 5 under conditions in which the nucleic acid molecule is expressed.

13. The isolated polypeptide of claim 12 comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number ____, an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number ____, or an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number ____.

14. A method for detecting the presence of a polypeptide of claim 8 in a sample, comprising:

a) contacting the sample with a compound which selectively binds to a polypeptide of claim 8; and

b) determining whether the compound binds to the polypeptide in the sample.

15. The method of claim 14, wherein the compound which binds to the polypeptide is an antibody.

16. A kit comprising a compound which selectively binds to a polypeptide of claim 8 and instructions for use.

17. A method for detecting the presence of a nucleic acid molecule of claim 1 in a sample, comprising the steps of:

a) contacting the sample with a nucleic acid probe or primer which selectively hybridizes to the nucleic acid molecule; and

b) determining whether the nucleic acid probe or primer binds to a nucleic acid molecule in the sample.

18. The method of claim 17, wherein the sample comprises mRNA molecules and is contacted with a nucleic acid probe.

19. A kit comprising a compound which selectively hybridizes to a nucleic acid molecule of claim 1 and instructions for use.

20. A method for identifying a compound which binds to a polypeptide of claim 8 comprising the steps of:

a) contacting a polypeptide, or a cell expressing a polypeptide of claim 8 with a test compound; and

b) determining whether the polypeptide binds to the test compound.

21. The method of claim 20, wherein the binding of the test compound to the polypeptide is detected by a method selected from the group consisting of:

a) detection of binding by direct detecting of test compound/polypeptide binding;

b) detection of binding using a competition binding assay;

c) detection of binding using an assay for Tango-72-mediated signal transduction.

22. A method for modulating the activity of a polypeptide of claim 8 comprising contacting a polypeptide or a cell expressing a polypeptide of claim 8 with a compound which binds to the polypeptide in a

sufficient concentration to modulate the activity of the polypeptide.

23. A method for identifying a compound which modulates the activity of a polypeptide of claim 8, comprising:

a) contacting a polypeptide of claim 8 with a test compound; and

b) determining the effect of the test compound on the activity of the polypeptide to thereby identify a compound which modulates the activity of the polypeptide.

```
CCTCGAGGGGAGGGGGAGATGCAACTCATCACATTTTTACTGACTGTCCTCTGGCTGTGCAAGTTATCTTGGAAGGGGG

ACTGGAAGAACAGTAATTCGGAGTCTGGGCTTGGCAGTTGGGCAAATCCAGGTTTACTCTTGGCTCTGCCACCTTCCAA

GAATGACACCTTGGTCAGATCTTTTAACCACACTGAGCCTCAGTTTTCCTCATCTCTAAAAGGGACTCGAAAATCTTAC

CAACTCATAGAGTTGGGGTGAGAATTCGAAGGTAATTCTATATAAGGTAAGGCCTCCAGCAAGAGCTATGGTGGTTGTG

ACACTGACTGAGGCTGGGGGAGGCCCTCACTCACCCTCCTTCCTTCTTGGTTTTCTCCTACCCAGATGTGGCAGTGGAT

GGGCAGAACTGGACGTTTGCTTTTGACTTCTCCTTCCTGAGCCAACAAGAGGATCTGGCATGGGCTGAGCTCCGGCTGC

AGCTGTCCAGCCCTGTGGACCTCCCCACTGAGGGCTCACTTGCCATTGAGATTTTCCACCAGCCAAAGCCCGACACAGA

                                        M   D   L   F   T   V   T   L   S   Q   V     11
GCAGGCTTCAGACAGCTGCTTAGAGCGGTTTCAG ATG GAC CTA TTC ACT GTC ACT TTG TCC CAG GTC     33

 T   F   S   L   G   S   M   V   L   E   V   T   R   P   L   S   K   W   L   K     31
ACC TTT TCC TTG GGC AGC ATG GTT TTG GAG GTG ACC AGG CCT CTC TCC AAG TGG CTG AAG    93

 R   P   G   A   L   E   K   Q   M   S   R   V   A   G   E   C   W   P   R   P     51
CGC CCT GGG GCC CTG GAG AAG CAG ATG TCC AGG GTA GCT GGA GAG TGC TGG CCG CGG CCC   153

 P   T   P   P   A   T   N   V   L   L   M   L   Y   S   N   L   S   Q   E   Q     71
CCC ACA CCG CCT GCC ACC AAT GTG CTC CTT ATG CTC TAC TCC AAC CTC TCG CAG GAG CAG   213

 R   Q   L   G   G   S   T   L   L   W   E   A   E   S   S   W   R   A   Q   E     91
AGG CAG CTG GGT GGG TCC ACC TTG CTG TGG GAA GCC GAG AGC TCC TGG CGG GCC CAG GAG   273

 G   Q   L   S   W   E   W   G   K   R   H   R   R   H   H   L   P   D   R   S    111
GGA CAG CTG TCC TGG GAG TGG GGC AAG AGG CAC CGT CGA CAT CAC TTG CCA GAC AGA AGT   333

 Q   L   C   R   K   V   K   F   Q   V   D   F   N   L   I   G   W   G   S   W    131
CAA CTG TGT CGG AAG GTC AAG TTC CAG GTG GAC TTC AAC CTG ATC GGA TGG GGC TCC TGG   393

 I   I   Y   P   K   Q   Y   N   A   Y   R   C   E   G   E   C   P   N   P   V    151
ATC ATC TAC CCC AAG CAG TAC AAC GCC TAT CGC TGT GAG GGC GAG TGT CCT AAT CCT GTT   453

 G   E   E   F   H   P   T   N   H   A   Y   I   Q   V   G   C   Q   A            169
GGG GAG GAG TTT CAT CCG ACC AAC CAT GCA TAC ATC CAG GTG GGA TGC CAG GCG           507

T                                                                                 508
```

FIG. 1

```
TTGGGACCCAGCAGGACACAGCAGTCAGGTGCATGCTGGGACCGCGACGGACAGGCTGCCGCACCCCAGGCCCCCA  79

                                                M  L  A  G  G  G  V  R          7
GAGGCCAGTCTGTTTGCCTCCCAACGCCATCTGACCCAGGTGAGCAAGAGG ATG CTG GCG GGG GGC GTG AGG 151

 S  M  P  S  P  L  L  A  C  W  Q  P  I  L  L  L  V  L  G  S                 27
AGC ATG CCC AGC CCC CTC CTG GCC TGC TGG CAG CCC ATC CTC CTG CTG GTG CTG GGC TCA 211

 V  L  S  G  S  A  T  G  C  P  P  R  C  E  C  S  A  Q  D  R                 47
GTG CTG TCA GGC TCG GCC ACG GGC TGC CCC CGC TGC GAG TGC TCC GCC CAG GAC CGC CGC 271

 A  V  L  C  H  R  R  K  R  F  V  A  V  P  E  G  I  P  T  E  T                 67
GCT GTG CTG TGC CAC CGC AAG CGC TTT GTG GCA GTC CCC GAG GGC ATC CCC ACC GAG ACG 331

 R  L  L  D  L  G  K  N  R  I  K  T  L  N  Q  D  E  F  A  S                 87
CGC CTG CTG GAC CTA GGC AAG AAC CGC ATC AAA ACG CTC AAC CAG GAC GAG TTC GCC AGC 391

 F  P  H  L  E  E  L  E  L  N  E  N  I  V  S  A  V  E  P  G                107
TTC CCG CAC CTG GAG GAG CTG GAG CTC AAC GAG AAC ATC GTG AGC GCC GTG GAG CCC GGC 451

 A  F  N  N  L  F  N  L  R  T  L  G  L  R  S  N  R  L  K  L                127
GCC TTC AAC AAC CTC TTC AAC CTC CGG ACG CTG GGT CTC CGC AGC AAC CGC CTG AAG CTC 511

 I  P  L  G  V  F  T  G  L  S  N  L  T  K  L  D  T  R  E  N                147
ATC CCG CTA GGC GTC TTC ACT GGC CTC AGC AAC CTG ACC AAG CTG GAC ACG AGG GAG AAC 571

 K  I  V  I  L  L  D  Y  M  F  Q  D  L  Y  N  L  K  S  L  E                167
AAG ATC GTT ATC CTA CTG GAC TAC ATG TTT CAG GAC CTG TAC AAC CTC AAG TCA CTG GAG 631

 V  G  D  N  D  L  V  Y  I  S  H  R  A  F  S  G  L  N  S  L                187
GTT GGC GAC AAT GAC CTC GTC TAC ATC TCT CAC CGC GCC TTC AGC GGC CTC AAC AGC CTG 691

 E  Q  L  T  L  E  K  C  N  L  T  S  I  P  T  E  A  L  S  H                207
GAG CAG CTG ACT CTG GAG AAA TGC AAC CTG ACC TCC ATC CCC ACC GAG GCG CTG TCC CAC 751

 L  H  G  L  I  V  L  R  L  R  H  L  N  I  N  A  I  R  D  Y                227
CTG CAC GGC CTC ATC GTC CTG AGG CTC CGG CAC CTC AAC ATC AAT GCC ATC CGG GAC TAC 811

 S  F  K  R  L  Y  R  L  K  V  L  E  I  S  H  W  P  Y  L  D                247
TCC TTC AAG AGG CTG TAC CGA CTC AAG GTC TTG GAG ATC TCC CAC TGG CCC TAC TTG GAC 871
```

FIG. 2 (1 of 3)

```
T  M  T  P  N  C  L  Y  G  L  N  L  T  S  L  S  I  T  H  C      267
ACC ATG ACA CCC AAC TGC CTC TAC GGC CTC AAC CTG ACG TCC CTG TCC ATC ACA CAC TGC  931

N  L  T  A  V  P  Y  L  A  V  R  H  L  V  Y  L  R  F  L  N      287
AAT CTG ACC GCT GTG CCC TAC CTG GCC GTC CGC CAC CTA GTC TAT CTC CGC TTC CTC AAC  991

L  S  Y  N  P  I  S  T  I  E  G  S  M  L  H  E  L  L  R  L      307
CTC TCC TAC AAC CCC ATC AGC ACC ATT GAG GGC TCC ATG TTG CAT GAG CTG CTC CGG CTG 1051

Q  E  I  Q  L  V  G  G  Q  L  A  V  V  E  P  Y  A  F  R  G      327
CAG GAG ATC CAG CTG GTG GGC GGG CAG CTG GCC GTG GTG GAG CCC TAT GCC TTC CGC GGC 1111

L  N  Y  L  R  V  L  N  V  S  G  N  Q  L  T  T  L  E  E  S      347
CTC AAC TAC CTG CGC GTG CTC AAT GTC TCT GGC AAC CAG CTG ACC ACA CTG GAG GAA TCA 1171

V  F  H  S  V  G  N  L  E  T  L  I  L  D  S  N  P  L  A  C      367
GTC TTC CAC TCG GTG GGC AAC CTG GAG ACA CTC ATC CTG GAC TCC AAC CCG CTG GCC TGC 1231

D  C  R  L  L  W  V  F  R  R  R  W  R  L  N  F  N  R  Q  Q      387
GAC TGT CGG CTC CTG TGG GTG TTC CGG CGC CGC TGG CGG CTC AAC TTC AAC CGG CAG CAG 1291

P  T  C  A  T  P  E  F  V  Q  G  K  E  F  K  D  F  P  D  V      407
CCC ACG TGC GCC ACG CCC GAG TTT GTC CAG GGC AAG GAG TTC AAG GAC TTC CCT GAT GTG 1351

L  L  P  N  Y  F  T  C  R  R  A  R  I  R  D  R  K  A  Q  Q      427
CTA CTG CCC AAC TAC TTC ACC TGC CGC CGC GCC CGC ATC CGG GAC CGC AAG GCC CAG CAG 1411

V  F  V  D  E  G  H  T  V  Q  F  V  C  R  A  D  G  D  P  P      447
GTG TTT GTG GAC GAG GGC CAC ACG GTG CAG TTT GTG TGC CGG GCC GAT GGC GAC CCG CCG 1471

P  A  I  L  W  L  S  P  R  K  H  L  V  S  A  K  S  N  G  R      467
CCC GCC ATC CTC TGG CTC TCA CCC CGA AAG CAC CTG GTC TCA GCC AAG AGC AAT GGG CGG 1531

L  T  V  F  P  D  G  T  L  E  V  R  Y  A  Q  V  Q  D  N  G      487
CTC ACA GTC TTC CCT GAT GGC ACG CTG GAG GTG CGC TAC GCC CAG GTA CAG GAC AAC GGC 1591

T  Y  L  C  I  A  A  N  A  G  G  N  D  S  M  P  A  H  L  H      507
ACG TAC CTG TGC ATC GCG GCC AAC GCG GGC GGC AAC GAC TCC ATG CCC GCC CAC CTG CAT 1651

V  R  S  Y  S  P  D  W  P  H  Q  P  N  K  T  F  A  F  I  S      527
GTG CGC AGC TAC TCG CCC GAC TGG CCC CAT CAG CCC AAC AAG ACC TTC GCT TTC ATC TCC 1711
```

FIG. 2 (2 of 3)

```
 N  Q  P  G  E  G  E  A  N  S  T  R  A  T  V  P  F  P  F  D    547
AAC CAG CCG GGC GAG GGA GAG GCC AAC AGC ACC CGC GCC ACT GTG CCT TTC CCC TTC GAC 1771

 I  K  T  L  I  I  A  T  T  M  G  F  I  S  F  L  G  V  V  L    567
ATC AAG ACC CTC ATC ATC GCC ACC ACC ATG GGC TTC ATC TCT TTC CTG GGC GTC GTC CTC 1831

 F  C  L  V  L  L  F  L  W  S  R  G  G  K  G  N  T  K  H  N  I    587
TTC TGC CTG GTG CTG CTG TTT CTC TGG AGC CGG GGC AAG GGC AAC ACA AAG CAC AAC ATC 1891

 E  I  E  Y  V  P  R  K  S  D  A  G  I  S  S  A  D  A  P  R    607
GAG ATC GAG TAT GTG CCC CGA AAG TCG GAC GCA GGC ATC AGC TCC GCC GAC GCG CCC CGC 1951

 K  F  N  M  K  M  I  *                                        615
AAG TTC AAC ATG AAG ATG ATA TGA                                1975

GGCCGGGGCGGGGGGCAGGGACCCCCGGGCCGGGCAGGGGAAGGGGCCTGGCCGCCACCTGCTCACTCTCCAGTCC 2054

TTCCCACCTCCTCCCTACCCTTCTACACACGTTCTCTTTCTCCCCTCCCGCCTCCGTCCCCTGCTGCCCCCCGCCAGCC 2133

CTCACCACCTGCCCTCCTTCTACCAGGACCTCAGAAGCCCAGACCTGGGGACCCCACCTACACAGGGGCATTGACAGAC 2212

TGGAGTTTAAAGCCGACGAACCGACACGCCGGCAGAGTCAATAATTCAATAAAAAGTTACGAACTTTCTCTGTAACTTG 2291

GGTTTCAATAATTATGGATTTTTATGAAAACTTGAAATAAAAAAAAAAAAAAAAAAAAG                    2351
```

FIG. 2 (3 of 3)

```
                                                  M  A  G  S  P          5
GAATTCGGCACGAGGCCAGCCAGTCCGCCSGYMCGRRGCCCGGCTCGCTGGGGCAGC ATG GCG GGG TCG CCG      72

 L  L  W  G  P  R  A  G  G  V  G  L  L  V  L  L  L  L  G  L          25
CTG CTC TGG GGG CCG CGG GCC GGG GGC GTC GGC CTT TTG GTG CTG CTG CTG CTC GGC CTG     132

 F  R  P  P  P  A  L  C  A  R  P  V  K  E  P  R  G  L  S  A          45
TTT CGG CCG CCC CCC GCG CTC TGC GCG CGG CCG GTA AAG GAG CCC CGC GGC CTA AGC GCA     192

 A  S  P  P  L  A  E  T  G  A  P  R  R  F  R  R  S  V  P  R          65
GCG TCT CCG CCC TTG GCT GAG ACT GGC GCT CCT CGC CGC TTC CGG CGG TCA GTG CCC CGA     252

 G  E  A  A  G  A  V  Q  E  L  A  R  A  L  A  H  L  L  E  A          85
GGT GAG GCG GCG GGG GCG GTG CAG GAG CTG GCG CGG GCG CTG GCG CAT CTG CTG GAG GCC     312

 E  R  Q  E  R  A  R  A  E  A  Q  E  A  E  D  Q  Q  A  R  V         105
GAA CGT CAG GAG CGG GCG CGG GCC GAG GCG CAG GAG GCT GAG GAT CAG CAG GCG CGC GTC     372

 L  A  Q  L  L  R  V  W  G  A  P  R  N  S  D  P  A  L  G  L         125
CTG GCG CAG CTG CTG CGC GTC TGG GGC GCC CCC CGC AAC TCT GAT CCG GCT CTC CGC GCC     432

 D  D  P  D  A  P  A  A  Q  L  A  R  A  L  L  R  A  A  R  L         145
GAC GAC CCC GAC GCG CCT GCA GCG CAG CTC GCT CGC GCT CTG CTC CGC GCC CGC CTT     492

 D  P  A  A  L  A  A  Q  L  V  P  A  P  V  P  A  A  A  L  R         165
GAC CCT GCC GCC CTA GCA GCC CAG CTT GTC CCC GCG CCC GTC CCC GCC GCG GCG CTC CGA     552

 P  R  P  P  V  Y  D  D  G  P  A  G  P  D  A  E  E  A  G  D         185
CCC CGG CCC CCG GTC TAC GAC GAC GGC CCC GCG GGC CCG GAT GCT GAG GAG GCA GGC GAC     612

 E  T  P  D  V  D  P  E  L  L  R  Y  L  L  G  R  I  L  A  G         205
GAG ACA CCC GAC GTG GAC CCC GAG CTG TTG AGG TAC TTG CTG GGA CGG ATT CTT GCG GGA     672

 S  A  D  S  E  G  V  A  A  P  R  R  L  R  R  A  A  D  H  D         225
AGC GCG GAC TCC GAG GGG GTG GCA GCC CCG CGC CGC CTC CGC CGT GCC GCC GAC CAC GAT     732

 V  G  S  E  L  P  P  E  G  V  L  G  A  L  L  R  V  K  R  L         245
GTG GGC TCT GAG CTG CCC CCT GAG GGC GTG CTG GGG GCG CTG CTG CGT GTG AAA CGC CTA     792

 E  T  P  A  P  Q  V  P  A  R  R  L  L  P  P  *                   261
GAG ACC CCG GCG CCC CAG GTG CCT GCA CGC CGC CTC TTG CCA CCC TGA                    840

GCACTGCCCGGATCCCGTGCACCCTGGGACCCAGAAGTGCCCCCGCCACCAGGACTGCTCCCCGCCAGCAC     919

GTCCAGAGCAACTTACCCCGGCCAGCCCTCTCACCCGAGGATCCCTACCCCCTGGC                    979
```

FIG. 3

```
T78          1 ..................................MDLFTVTLSQVTFSLG  16
                                                 |: |||. |||||. |
nodal      101 PMDIPTEGPLTIDIFHQAKGDPERDPADCLERIWMETFTVIPSQVTFASG 150

T78         17 SMVLEVTRPLSKWLKRPGALEKQMSRVAGECWPRPPTPP...ATNVLLML  63
                | |||||:|||||||| | |||||:|. |:.||..| |||    |.. .|||
nodal      151 STVLEVTKPLSKWLKDPRALEKQVSSRAEKCWHQPYTPPVPVASTNVLML 200

T78         64 YSNLSQEQRQLGGSTLLWEAESSWRAQEGQLSWE...WGKRHRRHHLPDR 110
               ||| .|||||||.|||||||||||||||||| |    ||:|:|||||||
nodal      201 YSNRPQEQRQLGGATLLWEAESSWRAQEGQLSVERGGWGRRQRRHHLPDR 250

T78        111 SQLCRKVKFQVDFNLIGWGSWIIYPKQYNAYRCEGECPNPVGEEFHPTNH 160
               |||||:||||||||||||||||||||||||||||||||||||||||||||
nodal      251 SQLCRRVKFQVDFNLIGWGSWIIYPKQYNAYRCEGECPNPVGEEFHPTNH 300

T78        161 AYIQV...GCQA..................................... 169
               ||||     :|:
nodal      301 AYIQSLLKRYQPHRVPSTCCAPVKTKPLSMLYVDNGRVLLEHHKDMIVEE 350
```

FIG. 4

```
T79          1 MLAGGVRSMPSPLLACWQPILLLVLGSVLSGS..ATGCPPRCECSAQDR.  47
                 : .:...   || |.:| |::.|: .:|  ..:||. | |. ..:
D45913       1 .....MARLSTGKAAC.QVVLGLLITSLTESSILTSECPQLCVCEIRPWF  44

T79         48 ..........AVLCHRKRFVAVPEGIPTETRLLDLGKNRIKTLNQDEFAS  87
                       .| |:  |:. :|:.:..:|.:| |..|.|
D45913      45 TPQSTYREATTVDCNDLRLTRIPGNLSSDTQVLLLQSNNI.........  84

T79         88 FPHLEELELNENIVSAVEPGAFNNLFNLRTLGLRSNRLKLIPLGVFTGLS 137
                          | ..:.:..|||| .|::. |.:. |. ..:..|.
D45913      85 ..............AKTVDELQQLFNLTELDFSQNNFTNIKEVGLANLT 119

T79        138 NLTKLDTRENKIVILLDYMFQDLYNLKSLEVGDNDLVYISHRAFSGLNSL 187
               .||.|.  ||.|  : || :||| ||..| :..|::  || .|||||..|
D45913     120 QLTTLHLEENQISEMTDYCLQDLSNLQELYINHNQISTISANAFSGLKNL 169

T79        188 EQLTLEKCNLTSIPTEALSHLHGLIVLRLRHLNINAIRDYSFKRLYRLKV 237
               .| |:. .|. |..  :.  ..| :|.: .  : :| | .|:.| .|:
D45913     170 LRLHLNSNKLKVIDSRWFDSTPNLEILMIGENPVIGILDMNFRPLSNLRS 219

T79        238 LEISHWPYLDTMTPNCLYGLN.LTSLSITHCNLTAVPYLAVRHLVYLRFL 286
               | :.   ||..:..|.| ||: |.|||: . .|. || ||:..:. |:||
D45913     220 LVLAG.MYLTDVPGNALVGLDSLESLSFYDNKLIKVPQLALQKVPNLKFL 268

T79        287 NLSYNPISTIEGSMLHELLRLQEIQLVG.GQLAVVEPY............. 323
               :|. ||| .|::: :.::|||.|:.: . |:|. |:.|
D45913     269 DLNKNPIHKIQEGDFKNMLRLKELGINNMGELVSVDRYALDNLPELTKLE 318

T79        324 .............AFRGLNYLRVLNVSGNQLTTLEESVFHSVGNLETLIL 360
                          |||::  |  | :..|.|..: :....|:.|| .: :
D45913     319 ATNNPKLSYIHRLAFRSVPALESLMLNNNALNAVYQKTVESLPNLREISI 368

T79        361 DSNPLACDCRLLWVFRRRWRLNFNRQQPT.CATPEFVQGKEFKDFPDVLL 409
               .|||| ||| : |: ..: .:.|  . .  || |.  .|.:.|   :||:
D45913     369 HSNPLRCDCVIHWINSNKTNIRFMEPLSMFCAMPPEYRGQQVK...EVLI 415

T79        410 PNYFT.CRRARIRDRKAQQVFVDEGHTVQFVCRADGDPPPAILWLSPRKH 458
               .:  . | .   :|  :.:: :| | |: : ||| ::|.|.|.|:.|  :
D45913     416 QDSSEQCLPMISHDTFPNHLNMDIGTTLFLDCRAMAEPEPEIYWVTPIGN 465

T79        459 LVSAKS.NGRLTVFPDGTLEVRYAQVQDNGTYLCIAANAGGNDSMPAHLH 507
                :.... .::..: .:|||||:   |::|.| | |:|.|..|.|...| :.
D45913     466 KITVETLSDKYKLSSEGTLEIANIQIEDSGRYTCVAQNVQGADTRVATIK 515

T79        508 V.....................................RSYSPDWPHQ 518
               |                                   :| .. . ..
D45913     516 VNGTLLDGAQVLKIYVKQTESHSILVSWKVNSNVMTSNLKWSSATMKIDN 565

T79        519 PNKTF............................AFISNQPGEGEANSTRA 540
               |: |:                          . | .|...: .| |
D45913     566 PHITYTARVPVDVHEYNLTHLQPSTDYEVCLTVSNIHQQTQKSCVNVTTK 615

T79        541 TVPFPFDIKTLIIATTMGFI..SFLGVVLFCLVLLFLWSRGKGNTKHNIE 588
               |.:|::||..  ..|.:: :  |:::|: :. : ::: .| | .. |:
D45913     616 TAAFALDISDHETSTALAAVMGSMFAVISLASIAIYIAKRFKRKNYHHSL 665

T79        589 IEYVPRKSDAGISSADAPRKFNMKMI........................ 614
               .|:.:.|. .:..   |. :|:.
D45913     666 KKYMQKTSSIPLNEL.YPPLINLWEADSDKDKDGSADTKPTQVDTSRSYY 714
```

FIG. 5

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/16241

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :C07H 21/02, 21/04, 1/00, 17/00; C12Q 1/68; G01N 33/53
US CL : 536/23.1; 530/350, 387.1; 435/6, 7.1
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 536/23.1; 530/350, 387.1; 435/6, 7.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

DIALOG: MEDLINE, BIOSIS, WPI, USPATFUL. author and terms (e.g. "TANGO" and protein) searched.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	Database MEDLINE on STN, 98072332, SONNENFELD, ET AL., The Drosophila tango gene encodes a bHLH-PAS protein that is othologous to mammalian Arnt and controls CNS midline and tracheal development. Development, November 1997, Volume 124, Number 22, pages 4571-82, Abstract.	1-23

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search	Date of mailing of the international search report
21 OCTOBER 1998	04 NOV 1998
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer HEATHER BAKALYAR Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*